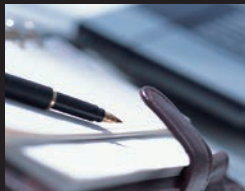
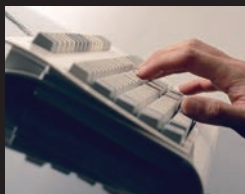
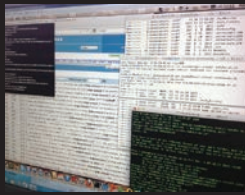


JAECS
Japan Association for English Corpus Studies

ISSN 1340-301 X

2021

英語コーパス学会



英語コーパス研究

第28号

English Corpus
Studies

28

2021年度 英語コーパス学会役員

会 長：石川慎一郎

副 会 長：家入 葉子（研究促進担当）
内田 諭（総務担当）
田畑 智司（学会誌担当）

事務局長：内田 諭

事務局員：阿部真理子、和泉 絵美（会計担当）
水本 篤、渡辺 拓人（広報担当）
三浦 愛香（総務担当）

理 事：アントニ・ローレンス、家入 葉子、石井 康毅、
石川慎一郎、石川 保茂、内田 諭、岡田 毅、
金澤 俊吾、小島ますみ、小林雄一郎、杉森 直樹、
住吉 誠、田畑 智司、塚本 聡、投野由紀夫、
水野 和穂、山崎 聡

幹 事：阿部真理子、五百蔵高浩、石川 有香、和泉 絵美、
今林 修、宇佐美裕子、大谷 直輝、加野まきみ、
谷 明信、長 加奈子、仁科 恭徳、能登原祥之、
藤原 康弘、三浦 愛香、水本 篤、渡辺 拓人

監 査：後藤 一章

顧 問：赤野 一郎、中村 純作、堀 正広

学会誌編集委員会

<委 員 長>能登原祥之

<委 員>今林 修、小島ますみ、能登原祥之、福元 広二、
水本 篤、田畑 智司

学会賞選考委員会

<委 員 長>塚本 聡

<委 員>石川 保茂、今林 修、高橋 薫、谷 明信、
西垣知佳子

大会実行委員会

<委 員 長>石川慎一郎

<委 員>和泉 絵美、内田 諭、杉森 直樹

ISSN 1340-301 X

英語コーパス研究

第 28 号

英語コーパス学会

2021

目 次

論文

Corpus-Assisted Discourse Studies in Airline Company Profiles Through the Lens of Moves and Adjectives Yasunori NISHINA	1
付加疑問との連鎖関係からみたテール（右方転位構造）の機能 — イギリス英語の会話コーパスを用いて — 山崎のぞみ	27

資料

英語コーパス学会第46回大会資料.....	47
-----------------------	----

「論文」

Corpus-Assisted Discourse Studies in Airline Company Profiles Through the Lens of Moves and Adjectives

Yasunori NISHINA

Abstract

This study quantitatively and qualitatively investigates the discourse of airline company profiles, using a DIY-tagged corpus of 7,735 words. It examines the move structure and language behavior in this specific discourse by investigating the subcorpora of the three alliances to which the 61 airline companies under investigation belong. Additional to the identification of move types and discourse structure, this research examines the two types of language units of adjectives—namely, colligations (i.e., ADJ + N) and semantic preferences (e.g., *best* + AWARD). Since company profiles comprise a representative genre of the business community, it is crucial to reveal this genre's shared and conventionalized knowledge and its culture to better understand the community.

1. Introduction

This study quantitatively and qualitatively examines the characteristics of 61 airlines' corporate profiles, using a DIY-tagged corpus of 7,735 words and identifying their discourse features (i.e., move types and typical move structures) and the adjectives' linguistic properties while focusing on key adjectives, colligations, and semantic preferences. Move analysis is an effective approach to discourse analysis in genre studies; it focuses on communicative functions and purposes referred to as *moves* (Bhatia, 1993; Swales, 1981, 1990). The current study uses move structures to investigate the rhetorical composition of texts within a genre.

The business discourse literature has focused on various practices, patterns, and similar language strategies in corporate narratives among companies. However, research has neglected affiliations such as alliances that intervene between companies

and discourses. This perspective is essential to a more precise analysis of corporate narratives. Thus, this study clarifies how the three alliances to which 61 companies belong affected the moves and language choice in the profiles, as such social factors interrelate with top-down language behavior and discourse structure. The literature confirms the close connections between the culture and values of a (discourse) community and the language used (e.g., Charles, 2004; Groom, 2005; Nishina, 2009). This study looks to reveal the shared knowledge and culture of airline industries.

2. Literature Review

2.1 The Concept of Genre

Genres constitute social phenomena and behavior (Bhatia, 1993; Mauranen, 1993; Swales, 1981, 1990). Each genre has specific communication purposes shared by the discourse community. Such purposes also affect the internal structure of genre texts. These internal structures often include typical keywords, collocations, colligations, semantic preferences, patterns, and/or semantic sequences showing particular communicative functions (Groom, 2005; Hunston, 2008; Ventola & Mauranen, 1996). Thus, genre can be used as a powerful lens through which to examine discourse and community culture logically in both academic and professional settings (Bhatia, 1993; Swales, 1981, 1990) and linguistically (e.g., conventionalized structure and linguistic units).

Indeed, the concept of genre and its relationship to discourse communities vary somewhat among the three main approaches of systemic functional linguistics (SFL), English for specific purposes (ESP), and the new rhetoric school. In particular, the SFL and ESP approaches have similarities in their conceptualization of genre, which is defined and/or restricted by formal textual features such as rhetorical structure and grammatical features (Bloor, 1998; Martin, 2003; Ventola, 1987). The current study also supports these genre approaches through analyses of language and discourse.

2.2 Corporate Narratives

Many researchers and practitioners in the business field have studied corporate narratives. Thomas (1997) examined the corporate narratives of letters in five years' worth of annual reports from a particular company; she found that the company

attempted to maintain a good impression and attract shareholders by strategically emphasizing the company's profitability. Leppanen (2012) also investigated the CEO letters of five Finnish companies to determine the language strategies used to rationalize their activities, and found that company management uses positive language to justify its activities. Thus, studies show that the corporate narratives genre is likely to present itself with confidence, optimism, and positive language.

Several other studies conclude that disclosure practices and patterns in corporate narratives differ among companies, although each narrative focuses on competitive advantages. These studies include Ocler (2009), who investigated the corporate social responsibility (CSR) reports of four French companies; Danilet and Mihai (2013), who investigated the online CSR discourse of three Romanian companies in the energy sector; and Hossain et al. (2016), who investigated the annual reports of Bangladeshi banking companies.

2.3 Corpus-assisted Discourse Analysis with a Small DIY Corpus

Corpus-assisted discourse analysis combines quantitative and qualitative discourse analysis to identify issues and how information is disclosed. Corpus-data use can overcome the shortcomings of the simple discourse analysis used in previous corporate narrative research (e.g., researchers' subjective biases and sample size limitations), as it narrows the target to a few subjects (in this case, airline companies) in each case study (e.g., four French companies [Ocler, 2009] and three NZ companies [Higgins and Walker, 2012], *inter alia*). Generalizing the research results and better understanding the language used in corporate narratives require analyzing a sufficient number of samples quantitatively and qualitatively. As Boulton (2012) points out, corpus linguistics is significantly better at pinpointing the conventionalized language used in a specific genre or text type, and such language regularities include discourse, collocations, lexical bundles, and keywords peculiar to a genre's texts.

Regarding corpus size for specialized texts, Bowker and Pearson (2002: p. 48) points out that "even corpora of between a few thousand and a few hundred thousand words have proved useful for language for special purposes (LSP) studies." In fact, various studies with small corpora have been conducted over several decades, since small DIY corpora are easy to compile and handle (Boulton, 2012). Aston (1997), for instance, compiled a small-sized corpus of 35,000 tokens from 12 medical articles,

while Curado Fuentes (2007) used a 40,000-word corpus of tourist advertisements. Focusing on corporate narratives, Hyland (1998) used a corpus linguistic approach to analyze CEO letters and director reports, compiling a DIY corpus from multiple annual reports. The current study clarifies the role that textual metadiscourse plays in assisting readers' understanding of textual structure, contextual information, and differences in the functions of metadiscourse among genres; it mirrors previous studies in its attempt to use a small DIY corpus to reveal the discourse features of a specific discourse.

2.4 Three Major Airline Alliances

This study targets company profiles for three main reasons. (1) In the study of corporate narratives, the company profile genre has been insufficiently studied. (2) It was easy to obtain all the companies' profiles and build the DIY corpus by restricting website searches to the airline industry (no random sampling was required). (3) Unlike in academic research, in business research, it can be challenging to distinguish between community-facing and industry-facing discourse; however, restricting both the industry focus (airlines) and the text source (public-facing company profiles) makes this possible.

Alliances provide benefits to airlines—such as pooling personnel, technology, and aircraft, developing route networks, and providing high standards of service and safety (He & Balmer, 2004). The three major airline alliances have different business aims and strengths. First, Star Alliance (SA), established in 1997, was formed by only five airlines; as of 2020, there were 26 members in this massive alliance, which is a leader in the airline business. Some key phrases are found on the alliance's website, such as “dedicated to innovation,” “excellent customer service,” “absolute reliability,” “smooth connections across a vast global network,” and “improve your travel experience” (<https://www.staralliance.com/>). Second, **oneworld** (OW) was founded in 1998 by five airlines; it now includes 13 members and provides service and connections to 1,000 destinations in 158 countries (N.B. “one” is officially indicated in boldface with lowercase as in **oneworld**). OW's slogan is “travel bright.” Once headquartered in Vancouver, Canada, the organization has been based in New York City since 2010. The motto found on the OW website is “make your flying experience as seamless as possible” (<https://www.oneworld.com/>). Finally, SkyTeam (ST) was founded in 2000 by four airlines and now includes 19 members, making it the second-largest airline

alliance in the world; it is headquartered in Amsterdam. Its network includes over 1,000 destinations in 175 countries. The catchphrases of this alliance are “make your travels smooth and enjoyable,” “optimizing existing services,” and “developing new benefits for our customers” (<https://www.skyteam.com/>).

All airlines ostensibly have the same goal: providing a comfortable flight experience for their customers. However, alliances differ in policies, membership requirements, customer loyalty strategies, and methods for improving SKYTRAX ratings. This study assumes that the external social factors surrounding these alliances influence the language used in their profile, and so corpus-assisted discourse analysis makes it possible to develop a detailed picture of discourse features and language behaviors. Thus, this study not only supports the validity of the move analysis of Swales (1981, 1990), Bhatia (1993), and others, but also shows that in typical business discourse, social factors have a significant influence on discourse content and structure.

3. Research Questions

Alliances have different goals, slogans, and histories that may affect discourse structure and language use in their corporate profile. However, if shared (language) knowledge and culture closely interrelate within the same genre and within a specific discourse community, it should be possible to identify similarities among the majority of profiles. These similarities pertain to common knowledge seen throughout the whole airline-company context, but dissimilarities nonetheless exist among alliances. Thus, this study looks to ascertain a common linguistic knowledge consistent throughout the genre of company profiles.

To accomplish this purpose, the current study combines move structure and adjective use analyses. Move structure analysis (e.g., Swales, 1990) is one of the most effective approaches in elucidating a specific discourse’s structure (e.g., Kondo, 2018). Adjectives are a key part of speech to be investigated, as their ratios and varieties are salient in written texts (Baker, 2003; Biber et al., 1999). For instance, Charles (2004) and Nishina (2009) elucidate the writer’s stance/evaluation system that is consistent in academic discourse, by investigating adjective patterns (e.g., it is ADJ that; ADJ PREP N) with corpus-based/driven approaches.

This study’s research questions are presented below. RQs 1–3 concern discourse

structure, RQ 4 concerns language behavior, and RQs 5–6 concern discourse community culture.

- (1) How many move types can be identified in airline company profiles?
- (2) Which moves are obligatory, conventional, and optional?
- (3) In airline company profiles, how is the typical move structure constructed?
- (4) What are the similarities among the three alliances in terms of adjective use?
- (5) What are the (dis-)similarities between this study and past corporate narrative studies?
- (6) What are the (dis-)similarities among the alliance profiles?

4. Analysis

4.1 Basic Corpus Data

The current study is based on a small corpus of airline company profiles manually compiled with CotEditor (ver.4.0.1) from the websites of three major alliances (i.e., SA, OW, and ST). When the corpus was compiled in 2019, there were 61 airline members, including 28 in SA, 13 in OW, and 20 in ST. Each alliance website presents a profile that includes information about its history, purpose, strengths, and other attributes. This study extracted from the websites only the company profiles and converted them into text files in a UTF-8 format; line breaks were manually erased.

Table 1 presents the basic airline company profile data, with 7,735 tokens and

Table 1. Basic data about airline company profiles

	Texts	Tokens (avg.)	Types (avg.)	TTR (STTR)	Sent (avg.)	Para (avg.)	AWL
SA	28	4,153 (148.32)	2,724 (97.29)	65.59 (66.89)	196 (7.00)	92 (3.29)	5.30
OW	13	684 (52.62)	577 (44.38)	84.36 (85.53)	33 (2.54)	15 (1.15)	5.12
ST	20	2,704 (135.20)	1,827 (91.35)	67.57 (68.90)	118 (5.90)	45 (2.25)	5.43
Company	61	7,735 (126.80)	5,183 (84.97)	67.01 (70.67)	346 (5.67)	151 (2.48)	5.16

Note: Sent = the number of sentences; Para = the number of paragraphs; AWL = average word length.

5,183 types.¹ CasualConc (ver.2.0.7) was used for the corpus analysis. Based on the average token, type, sentence, and paragraph scores, SA was found to have the most information per profile.

4.2 Move Analysis

Next, the sentence and paragraph positions of all the profiles were analyzed, and moves were identified through hint language expressions (e.g., lexico-grammatical patterns) (cf., Nishina, 2021). First, sample labeling was conducted, and an expert in the field validated the results. Finally, I checked and finalized the move analysis results. Table 2 shows the sample labeling at the initial stage of the move analysis, in which the profile of United Airlines included four distinct, tentative moves. Following the expert’s validation, I revised some of the moves (i.e., FOUNDATION→ALLIANCE HISTORY, STARTING COMPANY→FOUNDATION, CURRENT NETWORK→FLEET, NETWORK, CURRENT NETWORK→TIE-UP, OPERATION, NETWORK, EMPLOYMENT→EMPLOYMENT) (Table 2).

Table 2. Preparation for move analysis at the initial stage

Alliance	Airline	S	P	Example	Move
Star Alliance	United Airlines	1	1	United is a founding member of the Star Alliance network.	foundation
Star Alliance	United Airlines	1	2	The airline has a rich history in aviation, tracing its roots to 1926 when a small Swallow biplane owned by Walter T. Varney carried airmail to Nevada from Pasco, Washington – a flight that marked the true beginning of commercial air transportation and the birth of United Airlines.	starting company
Star Alliance	United Airlines	1	3	Today, the airline operates the most fuel efficient fleet among U.S. network carriers with the world’s most comprehensive global route network, including world-class international gateways to Asia and Australia, Europe, Latin America and the Middle East.	current network
Star Alliance	United Airlines	2	3	United, together with United Express, offers more than 4,500 flights a day to 339 destinations from hubs in Chicago, Denver, Guam, Houston, Los Angeles, New York, San Francisco, Tokyo and Washington, D.C.	current network
Star Alliance	United Airlines	1	4	United employs more than 87,500 people worldwide.	employment

In all, 333 sentences were manually counted; this number differs slightly from that in Table 1. As in Nishina (2021), the unit of a function was essentially identified at the clause level; in some cases, two distinct functions were found in a compound/complex sentence. In this study, however, several functions were sometimes found at a narrower level (e.g., the phrasal level). I thus identified each move based on functions found in a sentence, even when many moves were identified at the sentence/clause level. In the following example, the first move is “Its hub in Lisbon” (labeled as *BASE*); the second and third ones were *NETWORK* and *PRAISE*. Using this procedure, I pinpointed 18 types of moves (Table 3).

Its hub in Lisbon is a key European gateway at the crossroads of Africa, North, Central and South America, where TAP stands out as the international leading carrier in operation to Brazil. (TAP Air Portugal)

For instance, the move *OVERVIEW* concisely provides general introductory information about the company, especially in the first sentence of the first paragraph (45 instances; 51.72%). However, in many cases, the companies praised themselves subjectively through this move (76 instances; 87.36%), rather than provide simple information (11 instances; 12.64%). The following were the top five high-frequency modifiers used to emphasize company value: “leading” (14), “large” (14), “most” (11), “world” (11), and “flag(ship)” (9).

After revising the standard set of Amnuai and Wannaruk (2013),² I categorized moves with a range of 80 or more as obligatory, those with a range of 40–79 as conventional, and those with a range of under 40 as optional. Table 4 shows that in all alliances, the two moves *OVERVIEW* and *NETWORK* were obligatory; *ALLIANCE HISTORY*, *FOUNDATION*, and *FLEET* were conventional. The conditions for joining the alliance were that the airline company be large and financially strong, and operate international flights on a stable basis. Additionally, there had to be a positive factor for the other member companies (e.g., network). The purpose of a global airline alliance is to provide member-airlines’ customers with access to a vast global network of additional destinations, flights, and convenient connections; therefore, networks are crucial for any alliance (and for any airline). This is why the move *NETWORK* is considered obligatory overall (91.80%).

Table 3. The 18 moves in the airline company profiles

Move	Description	Example
OVERVIEW	This move describes the general introductory information about the company concisely.	<i>Aeroflot is Russia's de facto national carrier and largest airline...</i>
SOCIAL EVALUATION	This move includes awards, rankings, records, or stars that the company has received from third-party organizations, magazines, or surveys of professionals/the general public.	<i>ANA was voted Airline of the Year for 2013 by Air Transport World Magazines...</i>
ALLIANCE HISTORY	This move indicates when the company joined an alliance.	<i>EVA Air became a Star Alliance member on 18 June 2013...</i>
FOUNDATION	This move indicates information about the early history of a company and when and/or how the company was founded.	<i>Shenzhen Airlines Co., Ltd. was founded in November 1992 and started operations on September 17 of the following year...</i>
FLEET	This move indicates information about a company's aircraft fleet.	<i>It comprises 190 aircraft, which are modern Airbus A320-family, A330, Boeing 737, Boeing 777 and Sukhoi Superjet 100 airliners...</i>
NETWORK	This move indicates detailed information about national and international hubs, routes, and/or destinations.	<i>Aeroméxico, Mexico's global airline, serves more than 70 destinations in Mexico and major cities in the United States, Canada, Central & South America and Europe...</i>
STRATEGY	This move indicates the management strategy, investment, vision, slogans, and/or policies enacted/owned by the company.	<i>It plans to grow its fleet in line with its 10-Year Growth Strategy, increasing this to about 120 aircraft by 2021, and to expand its network to over 115 destinations during the same period...</i>
OPERATION	This move indicates information about a company's number of daily, weekly, or annual flights and/or the number of passengers carried.	<i>Currently with 8 US gateways, Turkish Airlines offers 70 flights per week, serving some 5,000 passengers...</i>

SERVICE	This move communicates information about various kinds of in/out-flight services devised and offered by the company.	<i>It has been recognized for its distinguished style of attentive, personalized onboard service (complimentary amenities, meals and beverages – including cocktails) in Coach and Clase Premier®...</i> <i>On 23rd October 2013 AEGEAN acquired Olympic Air...</i>
TIE-UP	This move refers to joint ventures with other (airline) companies, collaborations, or M&A.	
BASE	This move indicates the location of a company's headquarters, its main hub, or its home base including area, country, city name, or airport.	<i>Its operational headquarters is in Beijing, a major domestic and international hub in China...</i>
EMPLOYMENT	This move indicates the number of people employed by the company.	<i>United employs more than 87,500 people worldwide...</i>
ORGANIZATION	The move indicates the parent company, subsidiaries, or branch offices of a company.	<i>Czech Airlines is the subsidiary of Czech Aeroholding – a group of companies that operate in air transportation and related ground handling services...</i>
ENVIRONMENT	This move indicates how the company deals with issues related to the protection of the environment.	<i>TAROM is a modern and future-oriented airline, respecting the environment by implementing innovative procedures that reduce CO2 emissions...</i>
SAFETY	This move indicates how the company deals with issues related to the improvement of safety measures.	<i>For 56 years Thai Airways International has operated with high safety standards, adopting new safety regulations and practices as a prerequisite by various authorities around the world, and going even further to implement safety programs that are beyond the requirement...</i>
ASSOCIATION HISTORY	This move indicates when the airline company joined the International Air Transport Association (IATA).	<i>Aeroflot was the first Russian airline company to join the International Air Transport Association (IATA) in 1989...</i>
STOCK	This move indicates information about the stock listing of the company.	<i>On August 18, 2006, Air China was listed on Shanghai Stock Exchange under code 601111...</i>
WEBSITE	This move is likely to be added to a profile as a final remark to inform the audience of the location of further information.	<i>For more information about Middle East Airlines, Middle East Airline flights and partners, visit www.meu.com.lb...</i>

Table 4. Quantitative information about moves in the three subcorpora

moves	SA (184 sentences)			OW (33 sentences)			ST (116 sentences)			Total (333 sentences)		
	RF	Range (%)		RF	Range (%)		RF	Range (%)		RF	Range (%)	
OVERVIEW	22.08 (51)	85.71 (24)	obg	27.66 (13)	92.31 (12)	obg	14.94 (23)	80.00 (16)	obg	20.14 (87)	85.25 (52)	obg
SOCIAL EVAL	8.23 (19)	42.86 (12)	conv	4.26 (2)	15.38 (2)	opt	4.55 (7)	25.00 (5)	opt	6.48 (28)	31.15 (19)	opt
ALLIANCE HIST	3.03 (7)	25.00 (7)	opt	25.53 (12)	92.31 (12)	obg	3.90 (6)	30.00 (6)	opt	5.79 (25)	40.98 (25)	conv
FOUNDATION	6.93 (16)	46.43 (13)	conv	8.51 (4)	30.77 (4)	opt	5.19 (8)	40.00 (8)	conv	6.48 (28)	40.98 (25)	conv
FLEET	7.79 (18)	50.00 (14)	conv	4.26 (2)	15.38 (2)	opt	10.39 (16)	70.00 (14)	conv	8.33 (36)	49.18 (30)	conv
NETWORK	19.48 (45)	92.86 (26)	obg	27.66 (13)	92.31 (12)	obg	16.88 (26)	90.00 (18)	obg	19.44 (84)	91.80 (56)	obg
STRATEGY	6.06 (14)	32.14 (9)	opt	2.13 (1)	7.69 (1)	opt	7.79 (12)	60.00 (12)	conv	6.25 (27)	36.07 (22)	opt
OPERATION	4.33 (10)	35.71 (10)	opt				5.19 (8)	30.00 (6)	opt	4.17 (18)	26.23 (16)	opt
SERVICE	6.49 (15)	39.29 (11)	opt				11.04 (17)	55.00 (11)	conv	7.41 (32)	36.07 (22)	opt
TIE-UP	2.60 (6)	21.43 (6)	opt				5.19 (8)	20.00 (4)	opt	3.24 (14)	16.39 (10)	opt
BASE	4.76 (11)	28.57 (8)	opt				4.55 (7)	35.00 (7)	opt	4.17 (18)	24.59 (15)	opt
EMPLOYMENT	0.87 (2)	7.14 (2)	opt				0.65 (1)	5.00 (1)	opt	0.69 (3)	4.92 (3)	opt
ORGANIZATION	0.87 (2)	7.14 (2)	opt				3.25 (5)	20.00 (4)	opt	1.62 (7)	9.84 (6)	opt
ENVIRONMENT	0.43 (1)	3.57 (1)	opt				1.30 (2)	10.00 (2)	opt	0.69 (3)	4.92 (3)	opt
SAFETY	1.30 (3)	7.14 (2)	opt				2.60 (4)	15.00 (3)	opt	1.62 (7)	8.20 (5)	opt
ASSOCIATION HIST	0.87 (2)	7.14 (2)	opt				0.65 (1)	5.00 (1)	opt	0.69 (3)	4.92 (3)	opt
STOCK	1.30 (3)	7.14 (2)	opt				0.00 (0)	0.00 (0)	opt	0.69 (3)	3.28 (2)	opt
WEBSITE	2.60 (6)	21.43 (6)	opt				1.95 (3)	15.00 (3)	opt	2.08 (9)	14.75 (9)	opt
total	100 (231)	100 (28)		100 (47)	100 (13)		100 (154)	100 (20)		100 (432)	100 (61)	

Note: obligatory moves ≥ 80 ; 80 > conventional moves ≥ 40 ; 40 > optional moves

The SKYTRAX website explains ratings: “A typical rating of standards is based on analysis of between 500 and 800 product and service delivery assessment items. This covers airport services at the airline’s hub and onboard standards across all applicable cabin/aircraft types” (*About Airline Star Rating*, Skytaxratings.com, 2021).³ The SKYTRAX ranking of each airline is important in building customer loyalty. Keywords such as “service,” “hub,” and “cabin/aircraft” can be extracted from the website’s quoted information for use in airline company evaluations: “service” is included in the move SERVICE, “hub” in the move BASE, and “cabin/aircraft” in the move FLEET (Table 3). This is why the move FLEET is conventional overall (49.18%), especially for ST (70.00%) and SA (50.00%). The move SERVICE is also conventional for ST (55.00%) and optional, but almost conventional, in SA (39.29%). Since the move WEBSITE also indicates the IT service, its consideration changed the SA range scores; the total changed to conventional. The move BASE was treated as optional for SA (28.57%) and ST (35.00%); however, it showed the highest range score among the optional moves in ST. Therefore, SKYTRAX’s airline rating criteria and the alliance membership requirements affected the language used in the company profiles posted on each airline’s website.

Next, the move flow was calculated via [average scores of paragraph position (APP) \times average scores of sentence position (ASP)]. For example, in the case of the move FOUNDATION, the score was calculated as 1.32 (APP) \times 1.43 (ASP) = 1.89. By taking into account *both* the ASP and APP scores, it was possible to determine differences between two moves with similar ASP scores. Table 5 shows the order of the

Table 5. Move flow based on the APP \times ASP score

#	Move	APP	ASP	APP \times ASP	#	Move	APP	ASP	APP \times ASP
1	FOUNDATION	1.32	1.43	1.89	10	ORGANIZATION	2.14	3.71	7.94
2	ALLIANCE HISTORY	1.44	2.36	3.40	11	FLEET	2.06	3.94	8.12
3	OVERVIEW	1.74	2.57	4.47	12	SERVICE	2.16	4.41	9.53
4	ASSOCIATION HISTORY	2	2.67	5.34	13	EMPLOYMENT	2.67	4.33	11.56
5	OPERATION	1.89	3	5.67	14	SOCIAL EVALUATION	2.46	4.93	12.13
6	NETWORK	1.81	3.25	5.88	15	SAFETY	2.86	5.29	15.13
7	TIE-UP	1.86	3.57	6.64	16	STRATEGY	2.74	5.7	15.62
8	ENVIRONMENT	1.67	4.33	7.23	17	STOCK	2.33	7.33	17.08
9	BASE	2.11	3.67	7.74	18	WEBSITE	2.67	6.89	18.40

Table 6. Typical move structure

S	SP	P	Example	Move
1	1	1	Founded in 1950, Aerolíneas Argentinas is one of the leading South American carriers.	FOUNDATION, OVERVIEW
2	2	1	From its home bases in Aeroparque Jorge Newbery and Ezeiza International Airport in Buenos Aires, Aerolíneas Argentinas flies to 18 international destinations in The Americas, Europe and the South Pacific.	NETWORK
3	1	2	Along with Austral Líneas Aéreas, Aerolíneas operates flights to 35 destinations in Argentina, flying to more cities in the Argentine territory than any other airline.	NETWORK
4	2	2	Aerolíneas provides dynamic links between Argentina and the region through its relaunched hub in Aeroparque Jorge Newbery city airport.	NETWORK
5	3	2	The Aerolíneas group is carrying out an ambitious fleet renewal program.	STRATEGY

18 moves based on this score. Darkly shaded cells indicate obligatory moves, while lightly shaded ones indicate conventional moves, as in Table 4. Table 6 presents one sample by Aerolíneas Argentinas, an ST member, thus supporting the validity of this move structure.

4.3 Adjective Analyses

4.3.1 Data

The company profile corpus was annotated with the C7 tagset by the CLAWS part-of-speech tagger (<http://ucrel.lancs.ac.uk/claws7tags.html>). The line feed codes were then eliminated and converted into small spaces to make it possible to search collocations, n-grams, and colligations properly. Then, to capture the whole picture of the adjective distribution in the subcorpora of each alliance, the adjectives were extracted from the corpora by searching for JJ (general adjective), JJR (general comparative adjective) (e.g., “older,” “better,” “stronger”), JJT (general superlative adjective) (e.g., “oldest,” “best,” “strongest”), and JK (catenative adjective) (“able” in “be able to,” “willing” in “be willing to”).

Table 7 presents quantitative information about the types and tokens of adjectives, with their ratio in each of the three subcorpora; Table 8 presents the token ratio of adjectives in seven genres in the new WordbanksOnline.⁴ By comparing the data in Tables 7 and 8, we see that adjectives are more frequently used in company profiles than in the other genres, marking the highest token ratio (10.21%). Adjectives

constitute a key part of speech in airline company profiles.

Table 9 also shows the top 15 adjectives in each alliance, based on [relative frequency (RF) × file ratio (FR)] scores. This score quantitatively indicates each adjective’s importance within each alliance by considering the balance, frequency, and range of word choices. This table also presents adjective ratio (AR) information. Five adjectives—namely, *international*, *domestic*, *leading*, *best*, and *largest*—were ranked in all three subcorpora. Several adjectives are salient in the subcorpora of two alliances (i.e., SA–ST, SA–OW, SA–ST, SA–ST, and OW–ST); other adjectives are salient in individual alliances.

Table 7. Information types and tokens of adjectives in subcorpora

	SA	OW	ST	Company
Types (Ratio)	197(7.23%)	35(6.07%)	139(5.14%)	273(5.27%)
Tokens (Ratio)	447(10.76%)	56(8.19%)	287(10.61%)	790(10.21%)

Table 8. Token ratio of adjectives in seven genres from the new WordbanksOnline

Newspaper	6.64%	Broadcasts (radio & news)	7.09%	Ephemera (leaflets, newsletters, ads)	9.13%	News website	7.46%
Magazine	7.19%	Books (fiction & non-fiction)	6.94%	Informal speech	4.16%		

4.3.2 Adjectives featured in all alliances: General picture of airline company profiles

Next, it is important to consider findings from data on the colligation of adjectives, namely ADJ + N, pointing to specific semantic preferences,⁵ to uncover the features of specific discourses through the regularity of language use. As one of this study’s aims is to find consistent elements among company profiles, it should examine the adjectives that commonly occur in the subcorpora of two or three alliances. There were five adjectives common to all alliances—namely, “international,” “domestic,” “leading,” “best,” and “largest.” Table 10 summarizes the typical semantic preferences and collocates of these adjectives in the company profiles. In the example from Aegean Airlines (“AEGEAN, together with its subsidiary Olympic Air, provides scheduled passenger service directly to 145 destinations, 111 international, and 34 domestic, in 45 countries”), the adjective “domestic” modifies the noun collocate “destinations” at the L6 position, which cannot be identified through an automatic corpus search for the

Table 9. Adjectives in each alliance's profile in RF × FR score order

#	Adj. in SA	RF	AR(%)	FR(%)	RFxFR	Adj. in OW	RF	AR(%)	FR(%)	RFxFR	Adj. in ST	RF	AR(%)	FR(%)	RFxFR
1	<i>international</i>	6.46	6.71	75.00	484.58	<i>middle</i>	7.56	10.71	46.15	348.76	<i>modern</i>	3.98	4.18	60.00	238.86
2	<i>domestic</i>	3.02	3.13	39.29	118.46	<i>founding</i>	5.04	7.14	30.77	155.02	<i>international</i>	4.31	4.53	50.00	215.65
3	<i>leading</i>	2.80	2.91	39.29	110.01	<i>best</i>	2.52	3.57	15.38	38.74	<i>leading</i>	2.99	3.14	40.00	119.44
4	<i>best</i>	2.80	2.91	21.43	60.00	<i>domestic</i>	2.52	3.57	15.38	38.74	<i>largest</i>	2.99	3.14	30.00	89.58
5	<i>new</i>	2.59	2.68	21.43	55.40	<i>extensive</i>	2.52	3.57	15.38	38.74	<i>daily</i>	1.99	2.09	25.00	49.78
6	<i>regional</i>	2.15	2.24	25.00	53.85	<i>independent</i>	2.52	3.57	15.38	38.74	<i>middle</i>	2.99	3.14	15.00	44.79
7	<i>largest</i>	1.94	2.01	17.86	34.61	<i>international</i>	2.52	3.57	15.38	38.74	<i>other</i>	1.66	1.74	25.00	41.48
8	<i>central</i>	1.29	1.34	17.86	23.08	<i>largest</i>	2.52	3.57	15.38	38.74	<i>Atlantic</i>	1.99	2.09	20.00	39.82
9	<i>main</i>	1.29	1.34	17.86	23.08	<i>Latin</i>	2.52	3.57	15.38	38.74	<i>new</i>	1.66	1.74	20.00	33.18
10	<i>scheduled</i>	1.29	1.34	17.86	23.08	<i>leading</i>	2.52	3.57	15.38	38.74	<i>advanced</i>	1.33	1.39	20.00	26.54
11	<i>modern</i>	1.08	1.12	17.86	19.24	<i>main</i>	2.52	3.57	15.38	38.74	<i>best</i>	1.33	1.39	20.00	26.54
12	<i>other</i>	1.08	1.12	17.86	19.24	<i>major</i>	2.52	3.57	15.38	38.74	<i>domestic</i>	1.33	1.39	20.00	26.54
13	<i>efficient</i>	1.08	1.12	14.29	15.39	<i>recognised</i>	2.52	3.57	15.38	38.74	<i>European</i>	1.33	1.39	20.00	26.54
14	<i>swiss</i>	2.15	2.24	7.14	15.38	<i>star</i>	2.52	3.57	15.38	38.74	<i>innovative</i>	1.33	1.39	20.00	26.54
15	<i>haul</i>	1.29	1.34	10.71	13.84	<i>American</i>	2.52	3.57	7.69	19.37	<i>joint</i>	1.33	1.39	15.00	19.91

* Relative Frequency (RF) = word frequency / total tokens of each alliance * 1,000

* Adjective Ratio (AR) = adjectives / total adjectives * 100

* File Ratio (FR) = file range in each alliance

Table 10. Semantic preferences and collocates of the five adjectives common to all alliances

ADJ	Freq.	Typical Semantic Preference	Other Collocates
<i>international</i>	45	+COMPANY(11) [<i>airline(s)</i> (6), <i>carrier</i> (2), <i>airways</i> (1), <i>company</i> (1), <i>group</i> (1)]; +BASE(11) [<i>airport</i> (10), <i>hub</i> (1)]; +PLACE(7) [<i>destinations</i> (6), <i>cities</i> (1)]; +ROUTE(4) [<i>routes</i> (3), * <i>gateways</i> (1)]; +BUSINESS(2) [<i>logistics</i> (1), <i>trade</i> (1)] * <i>Gateways</i> are used <i>international gateways to ~</i> , indicating ROUTE rather than PLACE in the example.	<i>association</i> (4), <i>awards</i> (1), <i>brand</i> (1), <i>flights</i> (1), <i>passenger</i> (1), <i>presence</i> (1), <i>services</i> (1)
<i>domestic</i>	20	+COMPANY(7) [<i>airline</i> (3), <i>branches</i> (1), <i>carrier</i> (1), <i>group</i> (1), <i>offices</i> (1)]; +PLACE(5) [<i>destinations</i> (4), <i>cities</i> (1)]; +ROUTE(5) [<i>routes</i> (3), <i>network</i> (2)]	<i>hub</i> (1), <i>passenger</i> (1), <i>services</i> (1)
<i>leading</i>	28	+COMPANY(20) [<i>carrier(s)</i> (8), <i>airline(s)</i> (6), <i>group(s)</i> (3), <i>provider(s)</i> (2), <i>players</i> (1)]; +EVALUATOR(2) [<i>magazines</i> (1), <i>site</i> (1)]	<i>venture</i> (2), <i>brand</i> (1), <i>class</i> (1), <i>network</i> (1), <i>position</i> (1)
<i>best</i>	19	*+AWARD(16) [<i>airline(s)</i> (9), <i>service</i> (3), <i>dining</i> (1), <i>lounge</i> (1), <i>record</i> (1), <i>staff</i> (1)] *All SPs are double-quoted or followed by <i>award(s)/prize(s)</i> to indicate the name of award.	* <i>pilots/crews/staff</i> (1), <i>place</i> (1), <i>proof</i> (1) *Not indicating the name of award
<i>largest</i>	21	+COMPANY(19) [<i>airline(s)</i> (12), <i>carrier</i> (2), <i>group</i> (2), <i>operator</i> (2), <i>company</i> (1)]	<i>economies</i> (1), <i>part</i> (1)

colligation ADJ + N. Thus, I manually examined colligations and semantic preferences.

Airlines play an important role in connecting people within or among countries. It is thus naturally understood that the adjectives “international” and “domestic” would be key to all airline company profiles, irrespective of the alliances to which they belong. In terms of the adjective “international,” several airlines praise themselves as global companies, as represented by [*international* + COMPANY] (11/45; 24.44%). The nouns “airport” and “hub” are also salient collocates that point to another semantic preference [*international* + BASE] (11/45; 24.44%). The various global and local networks owned by airline companies are also featured in the following ways: [*international* + PLACE] (7/45; 15.56%), [*international* + ROUTE] (4/45; 8.89%), [*domestic* + PLACE] (5/20; 25%), and [*domestic* + ROUTE] (5/20; 25%). As Section 4.2 mentions, the semantic preference [ADJ + BASE/PLACE/ROUTE] reflects the fact that external social factors—such as SKYTRAX’s airline rating and alliance membership requirements— influence the language used in airline company profiles.

- Passenger traffic routes have reached 377, including 98 **international**, 16 regional, and 263 **domestic** routes. (Air China)
- Today, the airline operates the most fuel-efficient fleet among U.S. network carriers with the world’s most comprehensive global route network, including world-class **international** gateways to Asia and Australia, Europe, Latin America and the Middle East. (United Airlines)

Figure 1 contains word clouds of the collocates of *leading*, based on their raw frequency at the R1–R3 positions. The word cloud on the left comprises collocates, and that on the right comprises only nouns. The adjective “leading” occurs 28 times, often co-occurring with the nouns of the COMPANY category (e.g., “carrier(s),” “airline(s),” “group(s)” [20/28; 71.43%]). Owing to space limitations, I present as examples only the word clouds for “leading.”



Figure 1. Word clouds of collocates at R1–R3 of *leading* (left: any word; right: only nouns)

The general superlative adjective “best” occurs 19 times, and often co-occurs with the name of an AWARD received by the airline company to emphasize how superb it is (16/19; 84.21%) (e.g., *Business Traveler*’s “World’s Best Airline” Award, “Europe’s Best Airline,” and “Best Business Airline Lounge” prizes). Verbs such as “win,” “name,” “award,” “vote,” and “honor” also co-occur with [*best* + AWARD] units, sometimes in passivized forms. As a marker of AWARD, the units are often single- or double-quoted (11/19; 57.89%) without literally mentioning “award(s)” or “prize(s).” The generative superlative adjective “largest” also ranks in the top 15 adjectives among all alliances. This adjective is often used to praise the size, capacity, or ability of a

company in a country or specific area, and is followed by nouns such as “airline(s),” “carrier,” “company” to create the semantic preference [*largest* + COMPANY] (19/21; 90.48%).

- China Southern Airlines has been the **largest airline** in the People’s Republic of China for more than 35 years. (China Southern Airlines)

4.3.3 Adjectives featured in alliance pairs

The following five adjectives, all within the top 15, are salient among alliance pairs: “new” (SA–ST), “main” (SA–OW), “modern” (SA–ST), “other” (SA–ST), and “middle” (OW–ST). I manually counted the typical semantic preferences and their frequencies (Table 11).

Table 11. Semantic preferences and collocates of the five adjectives salient among alliances pairs

ADJ	Freq.	Typical Semantic Preference	Other Collocates
<i>new</i>	25	+PLACE(12) [<i>New York</i> (5), <i>New Europe</i> (3), <i>New Zealand</i> (3), <i>destinations</i> (1)]; +FLEET(7) [<i>aircraft</i> (2), <i>fleet</i> (2), <i>Airbus</i> (1), <i>Boeing</i> (1), <i>product</i> (1)]	<i>service</i> (2), <i>fares</i> (1), <i>shareholder</i> (1), <i>regulations and practices</i> (1), <i>Terminal</i> (1)
<i>main</i>	10	+BASE(9) [<i>hub</i> (9)]	<i>drive</i> (1)
<i>modern</i>	18	+FLEET(12) [<i>fleet(s)</i> (9), <i>aircraft</i> (2), <i>airliners</i> (1)]; +COMPANY(5) [<i>airline(s)</i> (4), <i>enterprise</i> (1)] * <i>Airliners</i> include <i>Airbus</i> , <i>Boeing</i> and <i>Superjet</i> in the example.	<i>technology</i> (1)
<i>other</i>	10	+COMPANY(6) [<i>airline(s)</i> (4), <i>carrier</i> (1), <i>subsidiaries</i> (1)]	<i>changes</i> (1), <i>hub</i> (1), <i>locations</i> (1), <i>service</i> (1)
<i>middle</i>	19	+DIRECTION(19) [<i>east</i> (18), <i>eastern</i> (1)] *Some instances seem ADJ+ADJ.	

* New Europe indicates central and eastern European countries that joined the EU in 2004.

“New” occurs more frequently in SA–ST because SA covers routes in most parts of Europe, the United States, and New Zealand. For instance, the national carrier of New Zealand, Air New Zealand, is an SA member; this fact reflects in the high frequency of area names including *New* in the semantic preference [*new* + PLACE] (19/25; 76%). The description of modern aircraft using [*new* + PLANE] (7/25; 28%) is

also salient. The adjective “main” is followed by the company’s home base—namely, [*main* + *BASE*] (9/10; 90%)—servicing national/international flights, essentially indicating large airports as its focus.

- [typical pattern: *main* + *hub* + *airport*] The airline and its affiliates service some 200 destinations in nearly 90 countries throughout Europe, North America, South America, Asia, Africa and Australia, with its **main hub London Heathrow**. (British Airways)

The adjective “modern” is used to emphasize the newness of aircraft operated by airlines (12/18; 66.67%) or the advanced vision held by companies (5/18; 27.78%). In other words, in their internal evaluations, two alliances (SA and ST) appear to hold the specific view that newness is an admirable value. These results imply the influence of external social factors on the language used in company profiles (as discussed in Sections 4.2 and 4.3.2) through the semantic preference [*ADJ* + *PLACE/FLEET/BASE/DIRECTION*], focusing on the global network, the hub airports, and cabin/aircraft.

5. Discussion

This study has shown that external conditions (e.g., SKYTRAX’s airline rating and alliance membership requirements) can influence the language airlines use in company profiles. It has also quantitatively shown how such language is used systematically in profiles. The findings support the validity of move analysis as proposed by Swales (1981, 1990), Bhatia (1993), and others, and shows that in business discourse, social factors can significantly influence discourse content and structure. Answers to the RQs follow.

(1) How many move types can be identified in airline company profiles?

- ▶ With the help of an expert in the field, this study identified 18 move types through three stages. See Table 3.

(2) Which moves are obligatory, conventional, and optional?

- ▶ The moves *OVERVIEW* and *NETWORK* are obligatory, and the moves *ALLIANCE HISTORY*,

FOUNDATION, and FLEET are conventional. The remaining 13 moves are optional. The purpose of a global airline alliance is to provide member-airline customers with access to vast global networks, and so networks are invaluable to alliances and member airlines. This is why the move NETWORK can be seen as obligatory. Additionally, each airline's SKYTRAX ranking is important in building customer loyalty: the keywords "cabin/aircraft" were extracted from the SKYTRAX evaluation and included in the move FLEET, explaining why that move is conventional.

(3) In airline company profiles, how is the typical move structure constructed?

► Moves are likely to follow a specific order, based on the scores of move discourse position: FOUNDATION → ALLIANCE HISTORY → **OVERVIEW** → (ASSOCIATION HISTORY) → (OPERATION) → **NETWORK** → (TIE-UP) → (ENVIRONMENT) → (BASE) → (ORGANIZATION) → FLEET → (SERVICE) → (EMPLOYMENT) → (SOCIAL EVALUATION) → (SAFETY) → (STRATEGY) → (STOCK) → (WEBSITE). (Boldfaced moves in this flow are obligatory, and moves in parentheses are optional.) These findings contribute to genre studies, since the move structures of airline company profiles have been previously unexamined.

(4) What are the similarities among the three alliances in terms of adjective use?

► Adjectives are more likely to be used in (airline) company profiles than in other profile types. Based on original RF × FR scores, it was found that several adjectives are commonly used by two or three alliances. An investigation of adjective colligation also showed that in this specific discourse, there are specific trends in the semantic preferences of high-demand adjectives.

(5) What are the (dis-)similarities between this study and past corporate narrative studies?

► This study showed that (airline) company profiles are no exception to the rule found by Thomas (1997) and Leppanen (2012) that positive language is likely to be used in corporate narratives (and business society) to justify past and current activities. Additionally, findings on corporate narratives (Danilet & Mihai, 2013; Hossain et al., 2016; Ocler, 2009) hold true for airline company profiles, in that disclosure practices and patterns in corporate narratives differ among airline companies and each profile is likely to focus on competitive advantages. The value of the current study is in its

finding that alliances also affect member-airlines' company profile language, as alliance policies differ (e.g., customer loyalty strategies, SKYTRAX ranking strategies, and membership conditions for joining). The examined discourse reflects this.

(6) What are the (dis-)similarities among the alliance profiles?

► The answer to this RQ is partly described in the previous answers. Some moves are commonly prioritized in all alliances, while others are featured in only some. A move is consistently structured, irrespective of alliance. The use of adjectives sometimes differs with the combination of two alliances. Tables 4 and 9 show that alliances differ in terms of their moves and adjectives. For instance, SA more heavily prioritizes international and domestic networks, relative to OW and ST.

6. Concluding Remarks

This study examined the discourse of airline company profiles through the lens of move structure and adjective usage, with the assistance of corpus-assisted methods; this method made it possible to solve the complex puzzle of this unknown discourse in a straightforward manner, using the moves, orders, scores, and language items detected herein. In line with many previous studies, this study found that each move includes a characteristic specific to a genre (Bhatia, 1993). Knowledge about move flow—specifically, the structural pattern of the text—is invaluable to understanding a specific genre. This study also discusses how external social factors implicitly restrict moves, move structures, and language use.

To resolve the problem of the size of the DIY corpus constructed herein—and support the results of this study—I would like in future research to compile a corpus of online annual reports. From the qualitative research, for example, I found that Singapore Airline's profile contents and its annual report correlate, and so its annual report can be considered a detailed version of its profile. Therefore, analysis of a company's annual report would clarify the specific content of the company's appeal, and the reliability and validity of the results would be enhanced on account of a larger dataset. However, since the annual report contains many images and infographics, qualitative discourse analysis may be more appropriate than a corpus analysis.

Under another agenda driven by the current study, I will examine in the near

future interrelations among various profile types of profiles, while focusing on their language behavior and discourse features.

Acknowledgements

I would like to express my gratitude to Prof. Judy Noguchi for validating the moves identified in this study. I would also like to thank three anonymous reviewers for their invaluable comments and suggestions. Needless to say, all remaining errors are my own.

Notes

1. As of 2019, the members of each alliance were as follows. [Star Alliance] (28 airlines): South African Airways, Air New Zealand, Brussels Airlines, Scandinavian Airlines, ANA, Air China, TAP Air Portugal, Croatia Airlines, Avianca, Asiana Airlines, Avianca Brasil, Swiss International Air Lines, Singapore Airlines, Austrian Airlines, Ethiopian Airlines, Shenzhen Airlines, Air Canada, Lufthansa, Eva Air, LOT Polish Airlines, Egyptair, Aegean Airlines, Air India, Adria Airways, United Airlines, Thai Airways, Copa Airlines, Turkish Airlines; [oneworld] (13 airlines): Qatar Airways, S7 Airlines, American Airlines, British Airways, LATAM, SriLankan Airlines, Royal Jordanian, Iberia, Japan Airlines, Malaysia Airlines, Finnair, Cathay Pacific, Qantas; [SkyTeam] (20 airlines): Aerolíneas Argentinas, TAROM, Aeroflot, China Airlines, Garuda Indonesia Airlines, Air Europa, Czech Airlines, Delta Air Lines, Alitalia, China Eastern Airlines, Vietnam Airlines, Xiamen Air, Air France, China Southern Airlines, Korean Air, Middle East Airlines, KLM, Saudia, Kenya Airways, Aeroméxico. Adria Airways and Avianca Brazil went bankrupt in September and October 2019.
2. According to Amnuai and Wannaruk (2013), moves occurring in every file are obligatory, those in the 60–99% range are conventional, and those below 60% are optional. However, the current study eased this standard, as it seemed too strict to label each move as obligatory, conventional, or optional. For instance, as per Amnuai and Wannaruk's (2013) standards, one move at the range 5% and the other move at the range 55% are both considered optional. The current study eased the ratio restrictions for obligatory and conventional moves to find the typicality of the discourse structure.
3. For more information, please visit the following websites:
 - (1) For the airlines' star ratings: <https://skytraxratings.com/about-airline-rating>.
 - (2) *SkyTeam News*: https://www.aeroflot.ru/us-ja/about/skyteam_alliance/skyteam_news/38289?_preferredLocale=us&_preferredLanguage=ja
 - (3) On alliances (joining, benefits, and pitfalls): http://www.airsource-partners.com/project/62-joining_an_alliance__benefits__pitfalls.html

- (4) Why do airlines join alliances? What are the benefits? at <https://simpleflying.com/airline-alliance-benefits/>
4. WordbanksOnline in the Shogakukan Corpus Network was renewed by the end of 2019. The corpus size has increased to 600 million through the addition of several new titles of magazines, reviews, novels, and new language data that mainly date from the 2010–2018 period. The search formula $\sim P(AJ.*)$ was used to extract the adjective ratio of each subgenre.
 5. According to Firth (1968), Sinclair (1996, 1998), Stubbs (2001), Tognini-Bonelli (2001), and Xiao and McEnery (2006), discourse features can be detected by investigating various conventionalized language units, such as collocations, colligations, semantic preferences, lexico-grammatical patterns, and lexical bundles; a colligation is the relationship between an individual word and grammatical categories (or originally between grammatical categories), and semantic preference is a semantic set of collocates.

References

- Aston, G. (1997) “Involving Learners in Developing Learning Methods: Exploiting Text Corpora in Self-access.” In Benson, P. and P. Voller (Eds.), *Autonomy and Independence in Language Learning*. London: Longman, pp. 204–214.
- Annuai, W. and A. Wannaruk (2013) “Investigating Move Structure of English Applied Linguistics Research Article Discussions Published in International and Thai Journals.” *English Language Teaching* 6: 1–13.
- Baker, M. (2003) *Lexical Categories: Verbs, Nouns, and Adjectives*. Cambridge: Cambridge University Press.
- Bhatia, V. K. (1993) *Analysing Genre: Language Use in Professional Settings*. London: Longman.
- Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan (1999) *Longman Grammar of Spoken and Written English*. London: Longman.
- Bloor, M. (1998) “English for Specific Purposes: The Preservation of the Species.” *English for Specific Purposes* 17: 47–66.
- Boulton, A. (2012) “Corpus Consultation for ESP: A Review of Empirical Research.” In Boulton, A, S. Carter-Thomas, and E. Rowley-Jolivet (Eds.), *Corpus-informed Research and Learning in ESP: Issues and Applications*. Amsterdam: John Benjamins, pp. 261–291.
- Bowker, L. and J. Pearson (2002) *Working with Specialist Language: A practical Guide to Using Corpora*. London: Routledge.
- Charles, M. (2004) *The Construction of Stance: A Corpus-based Investigation of Two Contrasting Disciplines*. PhD Dissertation, University of Birmingham, UK.
- Curado Fuentes, A. (2007) “A Corpus-based Assessment of Reading Comprehension in English.” In Hidalgo, E., L. Quereda and J. Santana (Eds), *Corpora in the Foreign*

- Language Classroom*. Amsterdam: Rodopi, pp. 309–326.
- Danilet, M. and O. Mihai (2013) “CSR Online Discourse Practices in the Romanian Energy Sector.” *Journal of Eastern Europe Research in Business and Economics 2013*: 1–9.
- Firth, J. R. (1968) *Selected Papers of J. R. Firth 1952–59*. London: Longman.
- Groom, N. (2005) “Pattern and Meaning Across Genres and Disciplines: An Exploratory Study.” *Journal of English for Academic Purposes 4*: 257–277.
- He, H. and J. M. T. Balmer (2004) “The Oneworld Alliance Brand: A Preliminary Inquiry.” *Working Paper Series 4*: 3–19.
- Hossain, D. M., A. T. S. Al Bir, K. M. Tarique, and A. Momen (2016) “Disclosure on Green Banking Issues in the Annual Reports: A Study on Bangladeshi Banks.” *Middle East Journal of Business 11*: 19–30.
- Hunston, S. (2008) “Starting with the Small Words: Patterns, Lexis and Semantic Sequences.” *International Journal of Corpus Linguistics 13*: 271–295.
- Hyland, K. (1998) “Exploring Corporate Rhetoric: Metadiscourse in the CEO’s Letter.” *Journal of Business Communication 35*: 224–245.
- Kondo, Y. (2018) “Move Development of London Hotel Overviews on Official Websites: Luxury Strategies in Overview Texts.” *English Corpus Studies 25*: 21–40.
- Leppanen, J. (2012) *Studying Corporate Texts: Critical Discourse Analysis of Finnish Companies’ Discourses on Growth in Russia*. Master’s thesis, Aalto University, Finland.
- Martin, P. (2003) “Genre and Discourse Community.” *ES: Revista de Filología Inglesa 25*: 153–166.
- Mauranen, A. (1993) *Cultural Differences in Academic Rhetoric*. Frankfurt am Main: Peter Lang.
- Nishina, Y. (2009) “A Study of ADJ PREP N Patterns in Two Soft-Applied Disciplines.” *Language Education & Technology 46*: 213–232.
- Nishina, Y. (2021) “Is an Impressive Background so Important to a CEO? Investigating the Discourse Structure of Personal Profiles in the Specific Business Field.” *LET Kansai Chapter Collected Papers 19*: 59–81.
- Ocler, R. (2009) “Discourse Analysis and Corporate Social Responsibility: A Qualitative Approach.” *Society and Business Review 4*: 175–186.
- Sinclair, J. (1996) “The Search for Units of Meaning.” *Textus IX*: 75–106.
- Sinclair, J. (1998) “The Lexical Item” In Weigand, E. (Ed.), *Contrastive Lexical Semantics*. Amsterdam: John Benjamins, pp. 1–24.
- Stubbs, M. (2001) *Words and Phrases*. Oxford: Blackwell.
- Swales, J. M. (1981) *Aspects of Article Introductions*. Birmingham: University of Aston.
- Swales, J. M. (1990) *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Thomas, J. (1997) “Discourse in the Marketplace: The Making of Meaning in Annual Reports.” *Journal of Business Communication 34*: 47–66.
- Tognini-Bonelli, E. (2001) *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

- Ventola, E. (1987) *The Structure of Social Interaction: A Systemic Approach to the Semiotics of Service Encounters*. London: Pinter.
- Ventola, E. and A. Mauranen (1996) *Academic Writing: Intercultural and Textual Issues*. Amsterdam: John Benjamins.
- Xiao, R. and T. McEnery (2006) “Collocation, Semantic Prosody, and Near Synonymy: A Cross-linguistic Perspective.” *Applied Linguistics* 27: 103–129.

(仁科 恭徳 神戸学院大学 E-mail: ynishina@gc.kobegakuin.ac.jp)

「論文」

付加疑問との連鎖関係からみたテール（右方転位構造）の機能 —イギリス英語の会話コーパスを用いて—*

山崎のぞみ

Abstract

The aim of the study is to explore how the co-occurrence of tails (right dislocations) with tag questions influences functions of tails performed in casual conversation. Tails often occur with tag questions in utterance-final position and can be placed either before or after them. Tails perform mainly two pragmatic functions: a disambiguating function for clarifying reference retrospectively, and a phatic and interpersonal function for creating effective connection with the hearer. Based on the Spoken British National Corpus 2014 (Spoken BNC2014), the data was analyzed with reference to (i) the order of tails and tag questions, (ii) the formal and meaning features of tails, and (iii) whether the utterance with a tail is evaluative or not. The analysis has shown that two types of tails exhibit different functional patterns according to the order with respect to tag questions. Noun phrase tails tend to serve to identify and clarify referents retrospectively when they follow tag questions, while they tend to perform the interpersonal function when preceding tag questions. In the case of pronoun tails, one-word tails, mostly *that*, almost always precede tag questions; but more than one-word tails are very likely to occur after tag questions, which suggests that they are closely related to the disambiguating function. It is shown that tails have their own combinatory behavior with respect to co-occurring tag questions.

1. はじめに

話し言葉に特徴的な言語形式の一つに、‘tail’と呼ばれる右方転位構造（right dislocation）がある（用語については2節で述べるが、以下「テール」と呼ぶ）。(1)の「あれは『バック・トゥ・ザ・フューチャー』にも出てくるよね、あの歌」という発話では、下線部の *that* と同一照応の要素 *that song* が再度、節の

域外に現れている。この *that song* のような言語要素をテールという。以下の発話例では、テールを太字で、テールが照応する節中の要素を下線部で表す。

- (1) S0653: that's in Back to the Future as well **that song** isn't it? (SZFG)¹

節を構成する統語要素として位置づけられない離節的要素であるテールは、伝統的な書き言葉中心の文法では、正規の語順から逸脱した周辺的あるいは方言的な非標準英語として、十分に扱われてこなかった。しかし、話し言葉研究の進展によって、この構造は話し言葉に万遍なく現れることが分かり、特定の変種や地域、社会階級に限定されない標準形式として認識されている (Carter, 1999: 154; Timmis, 2009: 343; Carter, Hughes, & McCarthy, 2011: 81; Durham, 2011: 257)。テールが現れる場所は、様々な語用論的調節が行われる発話の「周辺部」(澤田・小野寺・東泉, 2017) である。発話末要素の一種であるテールは、先行要素の指示対象の明確化や強調、話者の態度表明・感情表出、聞き手との連帯感の創出などの語用論的・談話的機能を持つことが指摘されている。

テールは、発話頭に起こる談話標識 (discourse marker, DM) が *and then* や *oh well* のようにしばしば連鎖して起こるのと同様、他の発話末要素と共起することが多い。*though*, *apparently* のような副詞句とも共起するが、(1) の *isn't it* のような付加疑問との共起が特に目立つ。評価や感情を緩和して伝えたり、聞き手を会話に巻き込んだりする付加疑問の機能はテールと共通する部分があるが、テールのふるまいや機能は単体で議論されることが多く、付加疑問と並置された場合の連鎖関係がテールの働きにどのように影響するのかという点については説明されていない。

本稿では、テールが付加疑問と共起する場合、選好される連鎖順序があるのかどうか、また、連鎖の仕方によってテールの機能に違いが見られるのかどうか、というリサーチクエスチョンに取り組む。イギリス英語の日常会話を収めた *The Spoken British National Corpus 2014* (*Spoken BNC2014*) を用いて、テールと付加疑問が共起している発話を収集し、テールの形式や意味、発話の評価的機能の面から分析する。本研究はテールと付加疑問の連鎖に限るが、これによって発話末についての理解、さらには発話末要素の連鎖構造についての理解を進めたい。

2. 研究の背景

2節では、テールの形式（2.1節）、テールの語用論的・相互行為的機能（2.2節）、テールが付加疑問など他の発話末要素と連鎖する場合（2.3節）についての先行研究を概観する。

2.1 テールの形式

テールは発話の周辺部である発話末に現れる。ここでは「発話」は基本的に、中心的メッセージ（発話の概念的核）（澤田・小野寺・東泉, 2017: 4）を有した統語的単位、あるいは、節のように統語的に完結していなくても一つの発話行為や命題を有した単位、さらに、その概念的核に付加された周辺部の要素から成る。(1)の発話は、以下のような構成で表され、テールの *that song* と付加疑問 *isn't it* が発話末要素である。発話の周辺部は、概念的核の前に置かれる発話頭もあり得るが、この発話には発話頭要素はない。

(1) S0653: that's in Back to the Future as well **that song** isn't it? (SZFG)
概念的核 | 周辺部（発話末）

テールの形式を指す名称や定義は様々である。Quirk et al. (1985: 1310, 1417) は ‘amplificatory tag’, Biber et al. (1999) は ‘noun phrase tag’, Leech (2000) は ‘end dislocation’, Durham (2011) は ‘right dislocation’ と呼んでいる。Aijmer (1989), Carter & McCarthy (1995, 2006), Rühlemann (2006, 2007, 2019), Timmis (2009, 2015a, 2015b) は、書き言葉が前提である左右への移動という空間的比喩を用いた ‘dislocation’（転位）ではなく、前後という時間的比喩の ‘tail’ という用語を使っており、これに倣って本稿でも「テール」と呼ぶ。

テールと見なされる形式も一通りではなく、反復や同格、省略、倒置などの言語現象と重なる側面を持つ。一般的にはテールは3～4タイプに分類される。最も標準的な形式は以下の (a) のような名詞句 (NP) テールで、次が (b) のような代名詞 (PRO) テールである。さらに、(c) のSV形式や倒置型の (d) VS形式のテールを含めることもある。

(a) S0167: he sounds like he's got some serious mental issues **that man** [NP]
(SAB7)

- (b) S0521: well that's interesting **that** [PRO] (SZNA)
 (c) S0611: it was bad maths **it was** [SV] (S2EF)
 (d) S0268: >>yeah it's a lot of **it is time** isn't it? [VS] (S69V)

この他, 'nice it was.' のように, テールが照応する要素が前の部分で省略されているものもある。照応要素の省略を伴うものを, Rühlemann (2007: 197) はテールの一形式と見なしているが, Durham (2011: 265) は, 倒置や補語の前置といった構造との区別が不明瞭のためテールには含めていない。このように, 用語の定義や含める形式は研究によって異なる。(c) (d) の SV / VS タイプは (a) (b) タイプより頻度が低いため, 本研究では, (a) の名詞句テールと (b) の代名詞テールのみを扱い, 照応要素の省略もテールに含める。

2.2 テールの機能

発話末は, 会話参加者のやりとりに重要な語用論的調節がなされる場所とされている(澤田・小野寺・東泉, 2017: 4)。Timmis (2009: 341) によると, 発話末のテールは, 会話の状況的要因である「リアルタイム性」(rea-time processing) と「話者同士の対人関係の調整」(relation management) に特に対応した言語現象である。テールが果たす語用論的・相互行為的機能を, この二つに沿って概観する。

第一に, 自然発生的な話し言葉に特徴的なテールは会話のリアルタイム性に即している。つまり, リアルタイムに会話を行う中で, ある要素が指示するものを後から思い出したように発話末で繰り返す。したがってテールは, 同じくリアルタイム性を要因とする会話の通常の非流暢性 (dysfluency) の特徴や, 修復 (repair), 反復 (repetition) といった言語現象と重なる面を持つ。(2) の会話例を見る。

- (2) S0264: erm and erm and the sandwiches were quite strange as well cos I I remember erm last last week there were erm I don't are they called shrimp or prawn?
 S0265: prawn probably
 S0264: and may- yeah sandwiches but erm I think they're quite expensive aren't they **prawns**? anyway there was nothing like that (SJAW)

話者 S0264は、下線部の they が指示するものが直前に言及した sandwiches ではなく prawns であることを、後から思い立って確認するようにテールで付け加えている。澤田（2016: 198）が「後出し的に指示を確定する」、Quirk et al. (1985: 1310) や Aijmer (1989: 148) が ‘postponed identification’ と言うように、テールには後付的に指示対象を確認・明確化する機能がある。(1) や (2) のように、テール（太文字）は名詞句で、テールが照応する要素（下線部）は代名詞の場合に特にその機能が見てとれる。さらに Biber et al. (1999: 958) がテールのことを、遡及的にトピックを標示する ‘retrospective topic markers’ と呼んでいるように、テールには談話で重要な要素を繰り返して強調する強意機能もある。

第二にテールには、「話者同士の対人関係の調整」に即して、聞き手との間に感情的結び付きや連帯感を作り出して心的距離を縮める対人関係的機能があると指摘されている（Aijmer, 1989; Carter & McCarthy, 1995; Rühlemann, 2007; Timmis, 2009; Carter, Hughes, & McCarthy, 2011; Durham, 2011）。テールを伴う発話は、会話参加者に既知の物事に対して、話者が即時的に態度や感情を表出する評価的な発話のことが多く、テールには聞き手との協調関係や共感をもたらす交感的機能があると認められている。Carter, Hughes, & McCarthy (2011: 81) は ‘Tails are an important feature of a listener-sensitive, affective grammar’ と評し、McCarthy & Carter (1997: 413) もテールを、共感や連帯感の創出を目的とした「やりとりのムーブ」(reciprocating move) の要素の一つに位置付けている。

Timmis (2009: 337) が述べているように、個々の文脈で使用されるテールは、指示対象の確認・明確化機能と対人関係的機能のどちらか一方だけを持つわけではなく、両方の機能を持つと考えられる。Biber et al. (1999) は指示対象の明確化をテールの主機能としている一方、McCarthy & Carter (1997) や Aijmer (1989) は対人関係的機能が主であると述べるなど、どちらの機能が優勢かという点については論じられている。また、テールの形式の違いを機能の違いに結び付けている研究もある。Rühlemann (2007: 198, 200) や Mycock (2017: 255) は、テールが名詞句の場合は指示対象の確認・明確化の機能を持つが、テールが代名詞の場合は付加疑問など他の発話末要素と共通する対人関係的機能を持つと述べている。

2.3 テールと他の発話末要素の連鎖

1節で指摘した通り、テールはしばしば他の発話末要素と共起する。発話頭

の連鎖については例えば, *and then* や *oh well* のような談話標識 (DM) の連鎖構造を論じた Lohmann & Koops (2016: 419) が, 類似の意味や機能の DM が共起しやすいという傾向を利用して個々の DM の意味や機能を判別することができる」と述べている。しかし, 発話頭の研究に比べて発話末要素の連鎖についての研究は進んでいない。テールに関しても同様だが, 一部の研究はテールが発話末で類似機能の語句と共起する点に言及している。例えば McCarthy (1998: 180) は, タグ, ヘッジ, モダリティ機能の語句, Carter, Hughes, & McCarthy (2011: 89) は *though* や *apparently* などの連結副詞や様態副詞と共起することを指摘している (*'It can lie dormant for years it can though apparently shingles'*)。本研究の対象である付加疑問との共起については, *'I just give it all away didn't I Rudy my knitting'* という例を挙げている Rühlemann (2007: 197) や, Mycock (2017: 265) が触れている。Carter, Hughes, & McCarthy (2000: 149) も, テールは付加疑問の前にも後にも起こり得るとして, 両者を使った会話を例示している。Timmis (2015b) は, 付加疑問との共起をテールの形式に関連づけて, 2.1節の4形式のうち, (a) (b) は付加疑問と共起するが, (c) (d) は共起しないと述べている。

上記のような共起情報は, それ自体, テールの働きや機能の識別のヒントになり得るが, 共起語と機能の関係をより明示した記述も散見される。Aijmer (1989: 151-152) は, *actually* や *really* などの副詞句が発話末でテールと共起することが, テールの社会的・対人関係的機能の証となり得ることや, 付加疑問などの緩和表現と共起するときはテールも緩和表現として機能することを記している。しかし以上のような研究でも, テールと他の発話末要素が共起する順序と機能の関係については扱われていない。

3. データと調査方法

本調査は The Spoken British National Corpus 2014 (Spoken BNC2014) というコーパスを利用する。Spoken BNC2014は, 2012~2016年に収集された668人のイギリス英語母語話者による1,150万語の会話コーパスで, 収められている会話は全てインフォーマルな日常会話である (Love et al., 2017)。自然発生的な会話コーパスとしては大きい規模だが, 2.1節で述べたように, テールの定義や形式は一通りではないため, 自動抽出や量的調査には困難が伴う。Timmis (2015b) が敢えて小規模コーパスを用いたように, 手作業による個別の判別が

避けられない。加えて本研究は、様々な形式があり得る付加疑問との共起を対象とするため、生起パターンも多岐に渡る。その結果、統計解析手法を用いた量的調査を実施できない難点があり、限られた数のデータを文脈的に分析する質的調査に重点を置かざるを得ない。しかし一方で、出現頻度が高いとは言えないテールと付加疑問の共起例を効率的に収集して傾向を見出すには、タグを利用した検索が可能なコーパス、それも、比較的大きい規模の会話コーパスが有用であると考え、Spoken BNC2014の利用を決定した。また、2000年代に話された日常会話を収める最新コーパスを用いることで、現在のイギリス英語における発話末要素の出現パターン的一端を捉えることができると考える。

したがって本研究では、コーパスからの網羅的抽出は目指さず、自動抽出と手作業による識別を繰り返して付加疑問の前後に現れるテールの収集を行った。付加疑問も形式が多様な上、テールよりはるかに高頻度のため、be動詞を含む肯定文の後に続く付加疑問（{be/V} n't _{PRON}）に限定した。ヒットする付加疑問の発話は進行形のものも含むが、テールを伴いやすい評価的・感情表出的なコンピュータ文の発話が比較的多いと予想されるためである。この付加疑問の検索式の左右に、名詞、代名詞、that（+名詞）、疑問符（?）、ターン末のタグ </u>² を様々なパターンで配して検索し、テールを含むものを手作業で抽出した。指示代名詞 that は、‘That was a crafty move **that**.’ (Rühlemann, 2007: 197) のように、名詞句テールにも代名詞テールにもしばしば用いられることが分かっている (Rühlemann, 2007: 197; Durham, 2011: 262; Timmis, 2015b: 319–320)。Spoken BNC2014は韻律情報を含まないため、注1にあるような最小限の音声情報は活用したが、テールかどうかの判断に迷うものは除外した。下記は、利用した検索式の数例である。that (_{N})? {be/V} n't _{PRON} / _{N} {be/V} n't _{PRON} \? </u> / {be/V} n't _{PRON} that / {be/V} n't _{PRON} that (_{N})? \?

4. 結果と分析

3節で述べた調査方法によって、表1の通り、付加疑問と共起するテールを名詞句 (NP) テール207例、代名詞 (PRO) テール82例、収集した。

表1 付加疑問と共起するテールのタイプ

	名詞句 (NP) テール	代名詞 (PRO) テール
収集数	207	82

表1のそれぞれのタイプのテールについて、付加疑問 (Tag Question, TQ) との順序を調べた結果が表2である。名詞句テールの合計数が表1と異なるのは、付加疑問がテールの前後、あるいはテールが付加疑問の前後についている場合は両方の順序にカウントしたためである。以下の分析・議論は、このデータを元に行う。

表2 テールと付加疑問 (TQ) の順序

	TQ + TAIL	TAIL + TQ
名詞句 (NP) テール (N=210)	165 (79%)	45 (21%)
代名詞 (PRO) テール (N=82)	39 (48%)	43 (52%)

4.1 名詞句テール

名詞句テールが照応する言語要素は (1) の下線部 *that* のように代名詞が最も一般的である (Timmis, 2009: 333)。本調査の名詞句テールも、207例のうち、先行する照応要素が代名詞のものが185 (89%) と圧倒的に多かった。したがって、照応要素より名詞句テールの情報量の方が多く、名詞句テールの主機能は指示対象の確認・明確化と推察される。以下では、付加疑問との共起を考慮に入れてテールの働きを考える。

4.1.1 テールと付加疑問の順序とテールの機能

Rühlemann (2007: 87) は、yes-no 疑問文の一形式である付加疑問の基本的機能を、‘to invite co-construction’, つまり、聞き手に会話の協同構築を働きかけることであると述べている。テールも聞き手志向の言語現象で、強調機能や聞き手との感情の共感や連帯感を作り出す機能は付加疑問と共通する。類似機能の両者の連鎖は、それらの機能をより強めていると考えられる。テールと付加疑問の共起関係をさらにつかむために、順序の関係を調べてみる。表2が示す通り、名詞句テールの場合、(3) のような「付加疑問+テール」(79%) の順序の方が、(4) のような「テール+付加疑問」(21%) より約4倍多い。

(3) S0136: >>yeah (.) so but it's a much better place now *isn't it?* **The park?**

(SC7H)

- (4) S0520: >>on in South Africa mm no that's right no I mean of course there's always an it's always a difficult thing **positive discrimination** *isn't it?*

(SQ3P)

この結果が示唆するところを考える。一般的に発話末は、次の話者に発言の順番が移るターン末となる可能性があり、ターンテイキングと関わるものが最後の方に置かれやすい傾向がある (Lohmann & Koops, 2016: 437)。Allerton (2009) が区別するように、付加疑問はその形式上、聞き手からフィードバックを引き出す機能を持つが、テールにはその機能はない。これらを考えると、聞き手へターンを明け渡す機能を持つ付加疑問の方がテールより後に置かれやすいのではないかと推察される。その逆の順序、つまり (3) のような、ターン末とより関係が深い付加疑問が先でテールが後という順序の場合、指示対象の明確化や強意のために後から付け加えられたというテールの「後付け感」が強まる。この順の方が高頻度ということは、名詞句テールの主な機能は指示対象の確認・明確化であることを示唆する。

次に、それぞれの順序のテールを、その発話が評価的（評価・価値判断・態度・感情を表す発話）か否かという点で分類した。分類に当たっては、Timmis (2015b: 317-318) が指摘するように、明らかに評価的な語句がない場合でも文脈によって態度や感情を表す発話もあるため、出来る限りコンテキストを広げて判断した。その結果を表3に示す。

表3 発話の評価機能と TQ, TAIL (NP) の順序

	TQ + TAIL	TAIL + TQ
評価的 (N=164)	127 (77%)	37 (23%)
非評価的 (N=46)	38 (83%)	8 (17%)

以下の (5) は TQ + TAIL の非評価的発話、(6) は TAIL + TQ の評価的発話の例である。

- (5) S0439: so where's the Lake District is below Manchester? (...) further?

S0441: up (.) it's in the Scottish border *isn't it* **Lake District?**

(S382)

(6) S0519: and they've rung up from the care people and said she needs to move
the bed away from the wall so she can get the frame in

S0520: right yeah that was always a tricky one **the bed** wasn't it? (STDN)

(5) のような非評価的な発話は、テールと共起する付加疑問は陳述の真偽の確認を求める機能を持つことが多い。共起するテールも、指示対象の確認・明確化機能が大きいと思われる。一方、(6) のような評価的な発話のテールは、評価の共感を求める対人関係の機能が強くなると推察され、付加疑問にも語調を和らげたり、聞き手に同意を促したりする機能が見られる。表3によると、評価的・非評価的発話とも、表2の名詞句テール一般の傾向と同じく、付加疑問の後にテールがくる TQ + TAIL の方が多いが、非評価的な発話における TQ + TAIL の割合 (83%) の方が、評価的な発話における割合 (77%) よりもやや高い。数が少なく頻度調査結果のみのため断定はできないが、(5) のような非評価的発話に多く見られる後付け的な TQ + TAIL のテールは、指示対象を確認・明確化する機能が大きく、(6) のような評価的発話に起こる傾向がやや高い TAIL + TQ は、聞き手との心的距離を縮める対人関係の機能が比較的大きいのではないかという仮説が立てられる。

4.1.2 テールの意味カテゴリー

4.1.1節で立てた仮説を裏付けるために、テールの名詞句が指すものを意味カテゴリーに分類した。分類は、名詞句の辞書的な意味のみによるのではなく、テールが照応する要素の指示対象をコンテキストから判断することによって決定した。

表4 名詞句の意味カテゴリーと TQ, TAIL (NP) の順序

	(具体的・物質的な) 物	事・概念	人	メタ言語	場所	動物
TQ + TAIL (N=165)	61 (37%)	42 (25.5%)	26 (15.8%)	10 (6.1%)	17 (10.3%)	9 (5.5%)
TAIL + TQ (N=45)	24 (53.3%)	3 (6.7%)	10 (22.2%)	2 (4.4%)	4 (8.9%)	2 (4.4%)

TQ + TAIL と TAIL + TQ の意味カテゴリーの割合で最も差が目立つのは、「事・概念」である。TQ + TAIL において「事・概念」を表す名詞句が25.5% と高く、

逆の順序における割合（6.7%）と差がある。TQ + TAIL の中でも「物」に次いで2番目の頻度を占める。(7)は「初めて住宅ローンを組むこと」という「事・概念」を表している（会話の表記方法は注1）。

- (7) S0192: this is it's mortgaged?
S0190: >>oh yes --UNCLEARWORD buying it yeah it's theirs
S0192: >>yeah so oh that's a big move *isn't it?* **getting your first mortgage**
(SW6A)

TQ + TAIL の「事・概念」のテールには、(7)のような動名詞や不定詞を用いた「～ということ（もの）」という説明的・概念的なテールがよく見られる。このことも、TQ + TAIL の順のテールは、前の代名詞の内容を後付け的に明確化する機能が大きいことを裏付ける。

一方、TAIL + TQ は、割合が一番高い「物」（53.3%）が、TQ + TAIL における割合（37%）よりも高く、また、「人」のカテゴリーの割合も同様である。(8)はTAIL + TQ のテールが「物」の例、(9)は「人」の例である。(8) (9)とも評価的発話で使われている。

- (8) S0571: --ANONnameM is gonna renew the team himself because they were
that good
S0517: >>no it's too expensive **that game** *isn't it?* (.) I've
S0572: >>>I don't wanna be in it this week
S0517: five pounds seventy something
(ST6R)
- (9) S0024: they were quite good **those golfers** *weren't they?* (S99N)

具体的・物質的な物や人を表すテールとTAIL + TQの順序の関係は、物や人に対する即時的な評価や感情を聞き手と共有しようとする対人関係的機能を示唆する。

さらに、表4のうちTQ + TAILとTAIL + TQの割合に差が見られる「物」「事・概念」「人」「メタ言語」の4カテゴリーのみを取り上げ、評価的・非評価的発話の分類を組み込んだものが表5である。

表5 名詞句の意味カテゴリー・発話の評価機能と TQ, TAIL (NP) の順序

	評価的 (N=140)		非評価的 (N=38)	
	TQ + TAIL	TAIL + TQ	TQ + TAIL	TAIL + TQ
(具体的・物質的な) 物	43	19	18	5
事・概念	38	3	4	0
人	21	9	5	1
メタ言語	5	2	5	0

表5によると、テールが表す名詞がメタ言語のものは12例あったが、2例を除いて全て TQ + TAIL の順である。メタ言語は言語そのものについて語るもので、テールは指示対象である言語を確認・明確化する機能が大きいと思われる。(10) は非評価的発話におけるメタ言語のテールの例である。

(10) S0617: that's how you say it *isn't it?* **Montenegro?** (ST82)

メタ言語のテールは、評価的・非評価的発話の両方において、TQ + TAIL の方が生起数が多く、この順序とテールの明確化機能の関係性が伺える。一方、表5の通り、TAIL + TQ で現れているメタ言語のテールは、2例とも評価的な発話であった。(11) はそのうちの1つである。

(11) S0496: George
 S0493: really?
 S0496: yeah
 S0493: quite like an old fashioned posh name **George** *isn't it?* (S6W8)

このテールにももちろん明確化機能はあるだろうが、評価的な陳述や主語と動詞の省略と相まって、評価や態度表明を聞き手と共有する対人関係的機能とより結び付いていると思われる。また、「事・概念」や「メタ言語」のテールは指示対象の明確化機能が大きいと指摘したが、この両カテゴリーの非評価的な発話では TAIL + TQ の順序がゼロだったことも、裏を返せば、TAIL + TQ が対人関係的機能の方とより関わっていることを示唆する。

また、上記の意味カテゴリーによる分類とは別に、(12) のような、人や物の名前、映画や歌の作品名などの固有名詞を使ったテールがどちらの順序で現

れているかを調べた。

(12) S0037: die Hard might might not be like a cinematic masterpiece

S0115: >>oh yeah I think that that was your favourite *wasn't it* **Die Hard?**

(S7MU)

表6 固有名詞の名詞句テール (N=46)

	固有名詞
TQ + TAIL (N=165)	41 (25%)
TAIL + TQ (N=45)	5 (11%)

表6は、それぞれの順序のテール全体の中で固有名詞が使われている割合を示しており、TQ + TAIL における割合の方が高いことが分かる。固有名詞のテールは、「メタ言語」や「事・概念」のテールと同様、指示対象の確認・明確化機能が大きいと考えられるので、やはりここでも TQ + TAIL の順序とこの機能の関連性が伺える。

4.1.3 指示代名詞 **that**

4.1.3節では、名詞句テールに用いられる指示代名詞の **that** を考慮に入れる。3節で言及した通り、テールには **that** がよく用いられる (Rühlemann, 2007: 197; Durham, 2011: 262; Timmis, 2015b: 320)。物理的に場面に存在する対象を指し示す直示用法の場合もあるが、基本的に、(13) のように会話に既に登場した旧情報を指示する前方照応用法である。

(13) S0663: so like black blue kind of grey like a slight?

S0661: >>ish (.) grey yeah it w- it was it was really nice actually **that colour**
wasn't it? (SXLC)

Rühlemann (2007: 197) や Durham (2011: 262) はさらに、**that** は話者の世界から聞き手の世界への視点の移動を表すため感情的含蓄を含み、表出された感情や態度を相手と共有しようとする対人関係的機能を持つと述べている。そのような機能を持つ **that** を含むテールと付加疑問の連鎖関係を調べるために、それぞれの順序に占める **that** / **those** を含むテールの割合を調べた。表7がその結

果である。

表7 that / those を含む名詞句テール (N=30)

	that / those を含む
TQ + TAIL (N=165)	17 (10%) (うち評価的発話16)
TAIL + TQ (N=45)	13 (29%) (うち評価的発話12)

that / those を含む名詞句テール30例のうち、ほとんどが評価的発話に現れており、視点を聞き手と共通の基盤へ移動させる that の特性と合致する。さらに、(13) のような TAIL + TQ のテールの方が that / those を含む確率がやや高く (29%)、このことも、この順序で現れるテールと対人関係的機能の関係を示唆している。

4.1.2, 4.1.3節の分析によって、4.1.1節で立てた仮説がある程度成り立つことが示された。つまり、発話末で付加疑問と共起する名詞句テールは、TQ + TAIL の順のテールは指示対象の確認・明確化機能を持つ傾向があるが、TAIL + TQ のテールは態度表明・感情表出を聞き手と共有する対人関係的機能が主であると考えられる。

4.2 代名詞テール

4.2節では、(14) のような代名詞テールについて考察する。

- (14) S0255: we walked there and walked all the way round and down by the river
and back round and long and
S0315: so it's a nice circuit **that** *isn't it?* (SGAN)

代名詞テールは名詞句テールと異なり命題の意味をほとんど含まないため、指示の繰り返しによる強意機能はあり得るが、指示対象の確認・明確化機能は考えにくい。代名詞テールの形式によって分類した結果は以下の通りである。

表8 代名詞テール (N=82) の形式

	that (all of that, that one 含)	those (those ones 含)	this (all this, this one 含)	they
生起数	72	6	3	1

Timmis (2015b: 319-320) や Mycock (2017: 258) などの先行研究が明らかにしているように、代名詞テールは *that* が圧倒的に多く、人称代名詞は稀である（本調査では *they* が一件のみ）。また、代名詞テールが照応する先行要素は、名詞句テールと同じく代名詞が69 (84%) と圧倒的に多く、(14) の *it* と *that* のように、節中でもテールでも名詞句による明示的指示がなされない場合がほとんどである。そのため Rühlemann (2007: 200) や Timmis (2015a: 96) は、代名詞テールの主機能を対人関係の機能と認定している。以下では、代名詞テールが付加疑問と共起する場合、4.1節の名詞句テールのように、連鎖構造によってテールの機能に違いが見られるのかどうかについて検討する。

代名詞テールと付加疑問の順序に関しては、表2が示す通り、TQ + TAIL と TAIL + TQ の割合はそれぞれ48%と52%で、名詞句テールの場合と異なって大きな差は見られなかった。この区別に、発話が評価的か非評価的かの区別を組み込んだのが表9である。ここでも、連鎖の順序に大きな割合の差は見られなかった。

表9 発話の評価機能と TQ, TAIL (PRO) の順序

	TQ + TAIL	TAIL + TQ
評価的 (N=70)	34 (49%)	36 (51%)
非評価的 (N=12)	5 (42%)	7 (58%)

次に、代名詞テールが *that* など1語の場合と、*that one*, *all of that* など2語以上から成る場合を比較した。その結果が表10である。

表10 代名詞テールの語数による分類と TQ, TAIL (PRO) の順序

	TQ + TAIL	TAIL + TQ
1語 (N=68) (<i>that</i> 61, <i>those</i> 5, <i>this</i> 1, <i>they</i> 1)	26 (38%)	42 (62%)
2語以上 (N=14) (<i>that one</i> 9, <i>this one</i> 1, <i>those ones</i> 1, <i>all of that</i> 1, <i>all this</i> 1, <i>that they had</i> 1)	13 (93%)	1 (7%)

数が少ないため一般化はできないが、2語以上の場合は1例を除いて全て、(15) のような TQ + TAIL の順序だった。

- (15) S0189: no I read that one
 S0192: you did?
 S0189: yeah
 S0192: do you like it?
 S0189: yeah it was good
 S0192: it's a good book *isn't it that one?* (S682)

本研究では all of that や that they had など代名詞テールに分類したが、表10に示した2語以上の代名詞表現の多くは、(15)のように物理的に存在する対象を直示する確認機能や all を用いた強調機能を持つ。このことは、より一層、TQ + TAIL 順で現れるテールと指示対象の確認・明確化機能の関連性の高さを示す。

一方1語のテールは、TAIL + TQ の割合 (62%) が、表2の代名詞テール全般における割合 (52%) より高い。1語の代名詞テールに限り、評価的発話と非評価的発話に分類した。

表11 代名詞テール (1語) の発話の評価機能と TQ, TAIL (PRO) の順序

	TQ + TAIL	TAIL + TQ
評価的 (N=58)	23 (40%)	35 (60%)
非評価的 (N=10)	3 (30%)	7 (70%)

評価的発話でも非評価的発話でも、TAIL + TQ の割合が、表2に示す代名詞テール全般における割合 (52%) より高いのは変わらない。(16) は評価的発話の TAIL + TQ の例である。

- (16) S0037: >>I watched that ages ago (.) there's some films that I really should've like I don't haven't seen many of the um Star Wars ones
 S0115: na (.) Bit overrated **that** *isn't it?* I think (S7MU)

(16) の評価的発話で相手に同意を求める付加疑問とともに使われたテール that は、4.1.3節で述べた that が持つ感情的含みも関係し、対人関係的機能が主と思われる。一方、非評価的発話の TAIL + TQ の that には、主に指示対象の確認や強意機能が見られた。つまり、1語の代名詞テールの場合、発話が評価的か否か、どちらの機能が主かということに関わらず、付加疑問の前に置か

れる TAIL + TQ の順序の傾向が高いことが分かる。

4.3 名詞句テールと代名詞テール

4.3節では、4.1、4.2節で分析・議論した結果に基づいて、名詞句テールと代名詞テールの2タイプのテールが持つ付加疑問との共起関係を比較・考察する。名詞句テールが発話末で付加疑問と共起する場合、連鎖順序によってテールの主機能が異なる傾向があり、TQ + TAIL のテールは指示対象の確認・明確化機能、TAIL + TQ は態度表明・感情表出を聞き手と共有する対人関係の機能が主となる。このことは、以下のような付加疑問の前後両方に現れているテールを比較するとよく分かる。

- (17) S0653: it does I mean it's it it is a really good book a really interesting book
and it goes through classic stories and films like style I mean that is
what George Lucas who wrote Star Wars he did use that book The
Hero with a Thousand Faces to base the st- the i- the plot and it's very
successful **that plot isn't it Star Wars?** (SCG9)

話者は、George Lucas が *The Hero with a Thousand Faces* という本を基にして *Star Wars* の筋を考えたという話をしている。発話末の付加疑問の前につけられた *that plot* は、直前の *the plot* の繰り返しであり、*very successful* という評価と相まって対人関係的なテールと考えられる。一方、付加疑問の後の *Star Wars* という固有名詞テールは、タイトル名が複数、言及されているので、どの話の何かを確認・明確化するためのテールであると言える。

一方、代名詞テールと付加疑問の連鎖関係は、テールの形式（語数）が関わっていると考えられる。*all* などを含む2語以上の代名詞テールは付加疑問の後に来る傾向があり、名詞句テールと同じく、TQ + TAIL のテールが強調や確認・明確化機能を持ちやすいことを裏付けた。一方 *that* のような1語テールの場合、TAIL + TQ の傾向がやや高く、発話が評価的か否かに関わらず、音韻的に短い1語のテールは概念的核の直後（発話末の始め）に添えられる傾向がある。1語の代名詞テールは、発話が評価的か否かに関係なく TAIL + TQ の順の割合が高いことから、名詞句テールで明らかになった TAIL + TQ の順と対人関係的機能の関連は代名詞テールでは不明であった。ただ、(18)のような代名詞テールと名詞句テールの共起例を見ると、評価的発話の *that* の対人関係的機能は明

らかである。

- (18) S0012: well that ain't no good is it? that's that's that's a bit lazy **that** *isn't it?*
just having a sandwich (SY4A)

(18) では、付加疑問の前に代名詞テール、後に名詞句テールを用いている。動名詞で表された TQ + TAIL の名詞句テールは明らかに指示対象の確認・明確化機能を持っており、一方で、前の代名詞テール *that* は話者の評価や感情表出と関わる対人的機能を担っていると見られる。

5. おわりに

本稿は、発話末要素であるテールの語用的機能を、付加疑問との共起という観点から論じた。発話頭の要素に比べて発話末要素の連鎖構造についての研究は進んでおらず、特にテールは単体で議論されることが多い。本研究は、テールがしばしば他の発話末要素、特に付加疑問と共起することに着目し、Spoken BNC2014を用いて、テールと付加疑問の連鎖関係がテールの働きにどのように影響しているのかということ明らかにした。

テールが付加疑問と共起する時、共起する順序によってテールが果たす主な機能に違いが見られ、名詞句テールと代名詞テールは、付加疑問との共起パターンが同じではないことが示された。名詞句テールが付加疑問と共起する場合、TQ + TAIL のテールは指示対象の確認・明確化機能、TAIL + TQ のテールは聞き手と感情や評価を共有する対人関係的機能が主となる傾向がある。一方、代名詞テールについては、all などを含む2語以上の場合、TQ + TAIL の順で現れて強調や確認・明確化機能を持つことが多いが、*that* のような1語テールは、付加疑問の前に来る TAIL + TQ の傾向がやや高い。代名詞テールの場合は、TAIL + TQ の順序と対人関係的機能の関連は不明であった。

本研究は、様々な語用論的調節がなされる発話末の解明には、どのような発話末要素が共起するかという点とともに、どのような順で共起するかという点に関わることを明らかにした。発話末の要素の連鎖が構造を持つことを示唆し、テールなどの発話末の研究には、他の発話末要素との共起や順序を考慮に入れる必要があることを示すものである。

注

* 本稿の内容は、英語コーパス学会第46回大会（2020年10月、オンライン）における研究発表に基づいている。執筆にあたり、3名の査読の先生方より貴重なご助言・ご指摘を数多く頂いた。ここに謝意を表したい。本稿における不備や誤りは全て筆者の責任である。なお、本研究は JSPS 科研費 16K02907 の助成を受けたものである。

1. 3節で述べる通り、本稿で引用する会話は The Spoken British National Corpus 2014 (Spoken BNC2014) からのものである。[S 番号] は話者 ID 番号を、引用末の括弧はテキスト ID 番号を示す。発話やターンの始まりは大文字ではなく、終わりもピリオドを用いていない。(.) は発話中に起こる1～5秒のポーズ、(...) は発話中やターン交替時の5秒を超えるポーズを示す。>> はそのターンの始まりが前のターンの終わり部分とオーバーラップしていることを示す。疑問符「?」は、形式的に明白な疑問や修辞疑問、付加疑問、明らかに上昇調で終わる陳述の発話につけられている。--ANONplace は非特定化された場所名、--ANONnameM と --ANONnameF は非特定化された男性名と女性名を表す。
2. ここでは、話者が話す順番を取ってから次の話者に発言の順番が移るまでを一つの「ターン」と考える。Spoken BNC2014でターン末を示すタグ </u> を利用すると、発話末のテールがターン末の場合のみを抽出する。2.2節の(2)のように発話末のテールの後もターンが維持される場合はヒットしないが、判別する件数を絞り込むために検索に用いた。

参考文献

- Aijmer, K. (1989) “Themes and Tails: The Discourse Functions of Dislocated Elements.” *Nordic Journal of Linguistics* 12: 137–154.
- Allerton, D. J. (2009) “Tag Questions.” In Rohdenburg, G. & Schlüter, J. (eds.), *One Language, Two Grammars?: Differences between British and American English*. Cambridge: Cambridge University Press, pp. 306–323.
- Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, S. (eds.) (1999) *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Carter, R. (1999) “Standard Grammars, Spoken Grammars: Some Educational Implications.” In Bex, T. & Watts, R. J. (eds.), *Standard English: The Widening Debate*. London: Routledge, pp. 149–166.
- Carter, R., Hughes, R., and McCarthy, M. (2000) *Exploring Grammar in Context: Grammar Reference and Practice, Upper-intermediate and Advanced*. Cambridge: Cambridge University Press.
- Carter, R., Hughes, R., and McCarthy, M. (2011) “Telling Tails: Grammar, the Spoken Language and Materials Development.” In Tomlinson, B. (ed.), *Materials Development in Language Teaching*. Cambridge: Cambridge University Press, pp. 78–100.
- Carter, R. and McCarthy, M. (1995) “Grammar and the Spoken Language.” *Applied Linguistics* 16, 2: 141–158.

- Carter, R. and McCarthy, M. (2006) *Cambridge Grammar of English: A Comprehensive Guide: Spoken and Written English Grammar and Usage*. Cambridge: Cambridge University Press.
- Durham, M. (2011) “Right Dislocation in Northern England: Frequency and Use — Perception Meets Reality.” *English World-Wide* 32, 3: 257-279.
- Leech, G. (2000) “Grammars of Spoken English: New Outcomes of Corpus-oriented Research.” *Language Learning* 50, 4: 675-724.
- Lohmann, A. and Koops, C. (2016) “Aspects of Discourse Marker Sequencing: Empirical Challenges and Theoretical Implications.” In Kaltenböck, G., Keizer, E., & Lohmann, A. (eds.), *Outside the Clause: Form and Function of Extra-clausal Constituents*. Amsterdam/Philadelphia: John Benjamins, pp. 417-445.
- Love, R., Demby, C., Hardie, A., Brezina, V., and McEnery, T. (2017) “The Spoken BNC2014: Designing and Building a Spoken Corpus of Everyday Conversations.” *International Journal of Corpus Linguistics* 22, 3: 319-344.
- McCarthy, M. (1998) *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.
- McCarthy, M. and Carter, R. (1997) “Grammar, Tails and Affect: Constructing Expressive Choices in Discourse.” *Text* 17, 3: 405-429.
- Mycroft, L. (2017) “Right-dislocated Pronouns in British English: The Form and Functions of ProTag Constructions.” *English Language and Linguistics* 23, 2: 253-275.
- Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. London: Longman.
- Rühlemann, C. (2006) “Coming to Terms with Conversational Grammar: ‘Dislocation’ and ‘Dysfluency’.” *International Journal of Corpus Linguistics* 11, 4: 385-409.
- Rühlemann, C. (2007) *Conversation in Context: A Corpus-driven Approach*. London: Continuum.
- Rühlemann, C. (2019) *Corpus Linguistics for Pragmatics: A Guide for Research*. London and New York: Routledge.
- 澤田茂保 (2016) 『ことばの実際1—話しことばの構造』東京：研究社。
- 澤田淳・小野寺典子・東泉裕子 (2017) 「周辺部研究の基礎知識」小野寺典子 (編) 『発話のはじめと終わり—語用論的調節のなされる場所』東京：ひつじ書房, pp. 3-51.
- Timmis, I. (2009) “‘Tails’ of Linguistic Survival.” *Applied Linguistics* 31, 3: 325-345.
- Timmis, I. (2015a) *Corpus Linguistics for ELT: Research and Practice*. London: Routledge.
- Timmis, I. (2015b) “Tails.” In Aijmer, K. & Rühlemann, C. (eds.), *Corpus Pragmatics: A Handbook*. Cambridge: Cambridge University Press, pp. 304-327.

英語コーパス学会 第46回大会

October 3 (Sat), 2020

Day 1 Schedule

Time	Event
08:00-09:00	Registration/Setup/Getting to know each other
09:00-09:30	Opening ceremony
09:30-10:30	Plenary 1: Jesse EGBERT (Northern Arizona University). ‘Corpus Linguistics’ or ‘Linguistics with a Corpus’?
10:30-12:00	Site exploration/Video viewing/Meeting others
12:00-13:00	Lunch break
13:00-13:25	<p>Presentation Q&A session 1 (JAECS members)</p> <p>Dax THOMAS (Meiji Gakuin University). Using the USAS Semantic Tagset To Explore Persuasive Language in Jeremy Taylor’s Holy Living and Holy Dying, 1650-1651</p> <p>Tatsuya ISHII (Kobe City College of Technology), Takeshi KAWAMOTO (Hiroshima University). Patterns Related To the Functions of Moves/Steps in the Introductions of Experimental Medical Research Articles: Combining Move Analysis and N-Grams Approach</p> <p>Laurence NEWBERY-PAYTON (Tokyo University of Foreign Studies), Sho FUKUDA (Toyama University), Keiko MOCHIZUKI (Tokyo University of Foreign Studies) . Learnability of English Verb-Particle Combinations and the Effect of Linguistically Motivated Instruction</p> <p>神澤 克徳 (京都工芸繊維大学), 小林雄一郎 (日本大学), 田中 悠介 (京都大学). 大学生を対象とする英語スピーキングテストの回答音声に基づくコーパス構築</p>
13:30-13:55	<p>Presentation Q&A session 2 (JAECS members)</p> <p>Akira MORIYA and Yoko IYEIRI (Kyoto University). The Be/Have-Perfect in 19th-Century American English: A Corpus-Based Analysis of Some Missionary Documents in Hawaii</p> <p>Risako AZEMOTO (Kyushu University). The Learners’ Development of Using Polysemous Words: The Case of ‘Over’</p> <p>仁科 恭徳 (神戸学院大学). ムーブと形容詞の振る舞いから見た航空会社プロファイルのディスコース分析</p> <p>投野由紀夫 (東京外国語大学)・川原田将之 (東京工業大学)・渡辺 亮嗣 (ネットアドバンス)・星野 守 (ネットアドバンス)・奥村 学 (東京工業大学). CEFR-J 準拠英語教育用 web コーパスの開発</p>
13:55-14:30	Break

Time	Event
14:30-15:30	<p>Presentation Q&A session 3 (JAECS non-members)</p> <p>Elen LE FOLL (Osnabrück University). Issues in Compiling and Exploiting Textbook Corpora</p> <p>Lukasz GRABOWSKI (University of Opole). Phrase Frames as an Exploratory Tool for Studying Translation Patterns: A Corpus-Based Descriptive Study</p> <p>Niall CURRY (Coventry University), Robbie LOVE (Aston University), Olivia GOODMAN (Cambridge University Press). Investigating Publisher Application of Corpus Research on Recent Language Change To ELT Coursebook Development</p> <p>Jamie WILLIAMS (Nottingham Trent University), David WRIGHT (Nottingham Trent University). Pronominal Ambiguity and Ascriptions of Responsibility in the UK Daily Coronavirus Briefings</p> <p>Ibrahim BASHIR (Jubail Industrial College), Kamariah YUNUS (Universiti Sultan Zainal Abidin). A Corpus Analysis of Prepositional Colligations in Nigerian Legal Discourse</p> <p>Barrios LEYRE (Universitat de Lleida), Vázquez GLÒRIA (Universitat de Lleida). Factuality and Conditional Sentences With Indicative Mode: A Corpus-Based Study</p> <p>Rosana VILLARES (University of Zaragoza). Corpus Linguistics Tools for the Creation of Linguistic Resources That Support the Internationalisation of Tertiary Education</p> <p>Hakan CANGIR (Ankara University), Taner CAN (TED University). Speaking of Extinction: A Comparative Corpus-assisted Analysis of the Environmental Framing in Climate Fictions and the News on the Net</p> <p>Yagang CHEN (The University of Edinburgh). Exploring the Use of Hedges in Academic Writing: A Corpus-Based Analysis Between Chinese TESOL Students and Expert Writers</p> <p>Xiaojing CHEN (The University of Edinburgh). A Comparative Study of Transition Markers in Thesis Abstracts From Chinese Undergraduates and RA Abstracts From Prestigious Journals</p> <p>Mat RAWSTHORNE (The University of Nottingham). Shared Experience: From I-Illness To We-Illness? Narrative Informed Corpus Linguistic Analysis of a Moderated Online Mood Disorders Forum</p> <p>Melissa KEMBLE (The University of Sydney). As Good as the Men? A Corpus-Based Analysis of Media Representations of Athletes Competing in the New Women's Australian Rules Football League</p>
15:30-16:00	Break
16:00-17:00	General Assembly
17:00-18:30	Welcome Reception (online)

October 4 (Sun), 2020**Day 2 Schedule**

Time	Event
08:00-08:45	Registration/Setup/Getting to know each other
08:45-09:00	Housekeeping/Notices
09:00-10:00	Plenary 2: Sowmya VAJJALA (National Research Council Canada). NLP Beyond NLPers – The Many Faces of NLP in Academia and Real-World
10:00-11:00	Site exploration/Video viewing/Meeting others
11:00-12:00	<p>Presentation Q&A session 4 (JAECS non-members)</p> <p>Martin SCHWEINBERGER (The University of Queensland). A Corpus-Based Analysis of Ongoing Change in the Adjective Amplifier Systems of Hong Kong and Philippine English</p> <p>Marine Laisa MATTE (Univates University), Larissa GOULART (Northern Arizona University), Simone SARMENTO (Federal University of Rio Grande do Sul), Rozane Rodrigues REBECHI (Federal University of Rio Grande do Sul). Becoming a President: A Diachronic Study on the Language of Brazilian President Jair Bolsonaro</p> <p>Yating YU (The Hong Kong Polytechnic University). Media Representations of ‘Leftover Women’ in China: A Corpus-Assisted Critical Discourse Analysis</p> <p>Michael HENSHAW (Hokkaido University). A Novel Approach To the ESP Keyword List: 2815 Entries With Frequent Lexical Bundles for Data-Driven Learning</p> <p>Will LINGLE (University of Aizu). Contrasting Narratives: The Greek Financial Crisis in Newspaper Editorials</p> <p>Wilfred Gabriel A. GAPAS (University of Santo Tomas), Rachele BALLESTEROS-LINTAO (University of Santo Tomas). The Discursive News Values of the 2017 Marawi City Crisis: A Corpus-Assisted Multimodal Discourse Analysis of Selected Newspaper Reports</p> <p>Ashleigh COX and Eric FRIGINAL (Georgia State University), Sabah S. MUSTAFA (University of Baghdad). Comparing Measures of Directness in Corpora of Essays Written by Iraqi EFL Learners, Native English-Speakers, and Advanced ESL College Students</p> <p>Andrew SCHNEIDER (Embry-Riddle Aeronautical University), Rachele UDELL (Georgia State University), Eric FRIGINAL (Georgia State University). Sky High: Building a Corpus of English for Flight Training</p> <p>Muchamad Sholakhuddin AI FAJRI (Universitas Airlangga). The Construction of Coronavirus in English-Language Indonesian Newspapers: A Corpus-Assisted Discourse Analysis</p> <p>寺田 里紗 (東京外国語大学). NICT JLE コーパスを用いた日本人英語学習者のイラスト描写における習得レベル別特徴分析</p>
12:00-13:00	Lunch break

Time	Event						
13:00-13:25	<p>Presentation Q&A session 5 (JAECS members)</p> <p>Reiko IKEO (Senshu University), Masayuki NAKAO (Tottori University), Eri SHIGEMATSU (Hiroshima University). A Corpus Stylistic Comparison of Speech Presentation of 21st-Century Present-Tense Fiction and 20th-Century Past-Tense Fiction</p> <p>Wei-Tung WANG (Tokyo University of Foreign Studies) and Yukio TONO (Tokyo University of Foreign Studies). Performance Evaluation of Automated CEFR Level Classification Tools</p> <p>Laurence ANTHONY (Waseda University), Natalie FINLAYSON, Emma MARSDEN, Rachel HAWKES, and Nick AVERY (National Centre of Excellence for Language Pedagogy, University of York). Extending Vocabulary Profiling To Languages Other Than English</p> <p>松田 佑治 (立命館大学). 現代英語における he/she of NP の振る舞い - COCA での調査から -</p>						
13:30-13:55	<p>Presentation Q&A session 6 (JAECS members)</p> <p>Hu XIAOLIN (Tokyo University of Foreign Studies). A Validation Study of the Accuracy of Lexical Diversity Tools</p> <p>Satoru UCHIDA, Takehiko SHIMIZU and Saaya KIMURA (Kyushu University). A Corpus Based Approach To Creating an Advanced Wordbook for University Students</p> <p>山崎のぞみ (関西外国語大学). 付加疑問との連鎖関係からみた右方転位構造 (テイル) の機能</p> <p>中谷 安男 (法政大学). 経済学国際ジャーナルにおける Method の章のコーパス分析</p>						
13:55-14:30	Break						
14:30-16:00	<p>Conference symposium/Discussion</p> <p>小・中・高における DDL 普及への挑戦 —DDL ツールの開発, 授業実践, 分野横断的考察—</p> <p>[Meeting the Challenges of Introducing DDL To Pre-Tertiary Students in Japan: Tool Development, Class Instruction and a Cross-Disciplinary Discussion]</p> <p>Part 1: Data-driven learning for younger learners: Current issues, future directions. Peter Crosthwaite (University of Queensland)</p> <p>Part 2: 小・中・高校生のための3種のウェブ DDL ツールと教材 :eDDL, hDDL, BES Search</p> <p>西垣知佳子 (千葉大学), 赤瀬川史郎 (Lago 言語研究所), 石井雄隆 (千葉大学), 神谷昇 (千葉大学). 西垣が発表</p> <p>Part 3: 小・中・高校における DDL 実践の分野横断的考察</p> <table border="0"> <tr> <td>①発見学習から DDL の効果を考える</td> <td>小山義徳 (千葉大学)</td> </tr> <tr> <td>②小学校英語教育の分野から</td> <td>物井尚子, 星野由子 (千葉大学)</td> </tr> <tr> <td>③小学校国語科教育の分野から</td> <td>安部朋世 (千葉大学)</td> </tr> </table> <p>Part 4: DDL の指導と評価</p> <p>水本 篤 (関西大学)</p> <p>Part 5: Communicative English Classroom using DDL in Thailand</p> <p>Pichinart Kumpawan (Surasakmontree School, Thailand)</p>	①発見学習から DDL の効果を考える	小山義徳 (千葉大学)	②小学校英語教育の分野から	物井尚子, 星野由子 (千葉大学)	③小学校国語科教育の分野から	安部朋世 (千葉大学)
①発見学習から DDL の効果を考える	小山義徳 (千葉大学)						
②小学校英語教育の分野から	物井尚子, 星野由子 (千葉大学)						
③小学校国語科教育の分野から	安部朋世 (千葉大学)						
16:00-16:15	Closing ceremony						

October 3 (Sat), 2020

< Plenary Talk 1 >

Plenary Talk 1: ‘Corpus Linguistics’ or ‘Linguistics with a Corpus’?

Jesse EGBERT

Associate Professor of Applied linguistics (Northern Arizona University)

Biography

Jesse Egbert is Associate Professor of Applied Linguistics at Northern Arizona University. Jesse specializes in register variation, quantitative methods in linguistics, and corpus linguistic approaches to legal interpretation. He is General Editor of the international peer reviewed journal *Register Studies*, and Technical Strand Editor for the series *Cambridge Elements in Corpus Linguistics*. He has published more than 60 peer-reviewed papers. Recent books include *Register Variation Online* (Cambridge, 2018), *Using Corpus Methods to Triangulate Linguistic Analysis* (Routledge, 2019), and *Doing Linguistics with a Corpus: Methodological Considerations for the Everyday User* (Cambridge, forthcoming).

Abstract

Corpus linguistics encompasses a vast array of empirical research. It appears that there is only one characteristic that holds true for all corpus linguistic studies: the use of a corpus. I propose that it is not only possible to use a corpus without doing linguistics; it is actually quite common, even in mainstream research publications. So what do I mean by using a corpus without doing linguistics? Linguistics can be defined simply as the scientific study of language. I propose that linguistics requires at least three conditions: (1) linguistically meaningful variables, (2) linguistically valid units of observation, and (3) linguistic description. In the absence of any one of these three conditions, (quantitative) corpus linguistics can quickly become nothing more than pattern hunting and number crunching. In essence, a computer can perform quantitative corpus analysis, but only a linguist can do linguistics with a corpus. Drawing on the research traditions of syntactic complexity, lexical dispersion, and keyword analysis, I will illustrate the importance of these three conditions by comparing studies that satisfy them to studies that don't.

< Presentation Q&A session 1 >

【研究発表1】

**Using the USAS Semantic Tagset to Explore Persuasive Language
in Jeremy Taylor’s Holy Living and Holy Dying, 1650-1651**

Dax THOMAS (Meiji Gakuin University)

Keywords: semantic tagging, persuasive language, Wmatrix

Abstract

This presentation reports on the initial stages of a study on persuasive language in two texts, *Holy Living and Holy Dying*, written by Jeremy Taylor in 1650 and 1651. The purpose of the study is two-fold: 1) to explore the persuasive language Taylor uses in his writing; and 2) to explore the usefulness of semantic tagging in this type of investigation. The corpus, consisting of the two Taylor texts (226,035 tokens), was first tagged with the USAS Semantic Tagset using the Wmatrix interface. A list of key concepts (semantic keyness) was generated using the Wmatrix interface and concordance lines were consulted for finer detail on the nature of the persuasive technique being used. Several persuasion techniques (such as “emotional appeal”, “attack”, “inclusive/exclusive language”) were selected and semantic tags were identified from the USAS tagset that related to each of these persuasion techniques. When exploring “emotional appeal”, for example, the “E” tag (Emotion) was used as a search item. While not all semantic tags resulted in useful search results, it was found that Taylor seemed to prefer negative persuasion techniques, such as appeals to fear and sadness in his writing. This is illustrated well by four out of the top five most frequent Emotion-related items being E4.1- (repentance), E5- (fear), E4.1- (sorrow), and E4.1- (sad). By working with a general-to-specific approach — that is, from persuasion technique, to general semantic concept category, to specific lexical item — elements of persuasion in the text could be readily identified.

【研究発表2】**Patterns Related To the Functions of Moves/Steps in the Introductions of Experimental Medical Research Articles: Combining Move Analysis and N-Grams Approach**

Tatsuya ISHII (Kobe City College of Technology)
Takeshi KAWAMOTO (Hiroshima University)

Keywords: move analysis, n-grams approach, keyword analysis, experimental medical research articles, introductions

Abstract

The aim of this presentation is to describe the patterns of prototypical three-grams to five-grams in the Introductions of experimental medical research articles (RAs) with the IMRD structure. Saber (2012), a pioneering study in the field of medical RAs, combined move analysis, and the n-grams approach. He built four subcorpora of moves based on IMRD sections to produce keyword lists. Moreover, he reported three-grams to five-grams with the keywords and identified the function of steps in each move. In this study, however, we generated 12 subcorpora based on the move analysis of IMRD to produce n-grams with stronger relevance to the functions of steps in each move. We collected 300 experimental medical RAs from 30 journals published in 2014 (approximately 1.5 million words in total).

Three moves that Nwogu (1997) described were found in Introductions: (Move1) Presenting the Background Information (40,412 words), (Move2) Reviewing the Related Research (80,875 words), and (Move3) Presenting the New Research (32,025 words). To identify the keywords in the three moves, we used CasualConc (Imao 2019) and calculated the scores of the Log-Likelihood Ratio. CasualConc produced three-grams to five-grams with the first keyword “is”, “been”, and “we” in Move 1, Move 2, and Move 3, respectively. We categorized these patterns into steps strongly associated with the functions of each move. The results of these patterns with the keyword show that Move 1 contains five patterns as Step 1 (established knowledge) and one pattern as Step 2 (problems). Move 2 includes three patterns as Step 1 (previous research) and two patterns as Step 2 (limitation). Move 3 involves three patterns as Step 1 (purpose), one pattern as Step 2 (procedure), and another single pattern as Step 3 (outcomes). We propose that corpora generation based on move analysis would lead to better pedagogically useful lists of academic phrases.

【研究発表3】

Learnability of English Verb-Particle Combinations and the Effect of Linguistically Motivated Instruction

Laurence NEWBERY-PAYTON (Tokyo University of Foreign Studies)

Sho FUKUDA (The University of Toyama)

Keiko MOCHIZUKI (Tokyo University of Foreign Studies)

Keywords: phrasal verb, verb-particle combination, learner corpus, SLA

Abstract

Verb-particle combinations (VPCs) have been shown to present difficulties for L2 learners (Liao & Fukuya 2004, Yoshitomi 2006, Siyanova & Schmitt 2007). Corpus studies have revealed that the difficulty of particular VPCs (Negishi, Tono & Fujita 2012) and their over- or under-use (Uchida 2012, Iio 2013, Ishii 2018) differ between L1 Japanese learners and native speakers or other learners. Acquisition studies (Yasuda 2010, Spring 2018, Nakagawa 2019) have assumed, explicitly or implicitly, that Japanese learners struggle to acquire VPCs since Japanese lacks such forms. Researchers have not attempted to explain VPCs with reference to forms in Japanese. This study examines the validity of such assumptions and the potential benefits of linguistically motivated instruction for Japanese learners. The research questions are as follows: 1. Is linguistically motivated instruction beneficial for VPC acquisition? 2. Does instruction making explicit reference to learners' L1 have particular benefits for VPC acquisition? Three sets of audio-visual resources were created to explain VPCs with a. insights from cognitive linguistics; b. insights from contrastive linguistics; c. a mix of the two. Four groups of 1st year non-English majors participated in the study. Three groups received one of the three audio audio-visual resources, while the fourth

served as a control group. All groups then memorized a list of VPCs. Results are reported for pre-, post- and delayed post-test scores. The group using the “mixed” resources showed the only significant increase in average score between pre- and delayed post-test results, suggesting the resources were effective at aiding VPC acquisition. This increase in average score was observed for 80% of learners, suggesting the wide applicability of the resources. Furthermore, the increase in scores was not restricted to items explicitly covered in the resources, suggesting learners became better able to infer the meanings of VPCs more generally.

【研究発表4】

大学生を対象とする英語スピーキングテストの回答音声に基づくコーパス構築

神澤 克徳（京都工芸繊維大学）

小林雄一郎（日本大学）

田中 悠介（京都大学）

Keywords: コーパス構築, 学習者話し言葉コーパス, 英語スピーキングテスト, 大学生

Abstract

本発表の目的は、発表者らが現在構築中のコーパスである KIT Speaking Test Corpus の概要を述べ、それを使用した分析例を紹介することである。KIT Speaking Test Corpus は、京都工芸繊維大学で1年次生全員を対象に実施した英語スピーキングテスト (KIT Speaking Test) の回答音声を書き起こしたものである。書き起こしの対象となる回答音声は74時間16分（1名あたり7分45秒×575名）である。書き起こしデータには「フィラー」や「自己訂正」など17種類のタグを挿入している。また、ヘッダー情報として、受験者の属性のほか、スピーキングテストのスコアや直前に受験した TOEIC のスコアを付与している。このような仕様によって、学習者の習熟度（テストスコア）とパフォーマンスの関係、スピーキングテストのタスクが受験者のパフォーマンスに与える影響などの分析が可能となり、さまざまな角度から日本語を母語とする大学生の英語スピーキング能力の実情を解明することができると期待される。本発表では、4段階の習熟度別の産出語彙量、特徴的な語句を報告する。具体的には、3人称代名詞、副詞、時制、接続表現、前置詞句、that 節などの発達パターンに注目する。また、フィラーや自己訂正の頻度と生起位置に注目することで、学習者が言いよどみやすい箇所を光を当てる。

< Presentation Q&A session 2 >

【研究発表 1】

**The Be/Have-Perfect in 19th-Century American English:
A Corpus-based Analysis of Some Missionary Documents in Hawaii**

Akira MORIYA (Kyoto University)

Yoko IYEIRI (Kyoto University)

Keywords: historical linguistics, English in Hawaii, perfect auxiliaries, Late Modern English**Abstract**

The present study discusses the be- and have-perfect in the ABCFM Hawaii Corpus (the Hawaii Corpus), a corpus of approximately 653,100 words compiled by our research team, using mainly journals written by eight members of the American Board of Commissioners for Foreign Missions (ABCFM), who migrated to Hawaii in the 19th century (cf. Iyeiri & Fukunaga, 2020, for some details of the corpus).

Both be- and have-auxiliaries are known to have been used in the perfect construction of mutative intransitive verbs such as come and go since the Old English period (e.g. ‘they are gone’ (be-perfect) and ‘they have gone’ (have-perfect), cf. Denison, 1998). In the Late Modern English period, however, the have-perfect began to replace the be-perfect, and became dominant as in contemporary English (Rydén & Brorström, 1987). This study aims to investigate the uses of perfect forms in the Hawaii Corpus, to compare them with the general trend in 19th-century English and to account for the characteristics of the be/have-perfect usage in different texts included in the Hawaii Corpus. Kytö (1997) suggests that American English (AmE) experienced an earlier increase in the rate of the have-perfect usage compared to British English, and the analysis of the Hawaii Corpus reveals indeed that the be-perfect is already quite rare, except for a few verbs such as improve and change, confirming the general trend in AmE. Other findings in the Hawaii Corpus include: the verb improve, showing a relatively frequent use of the be-perfect, prominently appears in certain contexts regarding health (e.g. her health is very much improved); the be-perfect is less frequent in journals than other text genres; and there are considerable differences in the preference of perfect auxiliaries even between a married couple.

【研究発表 2】

The Learners’ Development of Using Polysemous Words: The Case of ‘Over’

Risako AZEMOTO (Kyushu University)

Keywords: second language acquisition, polysemous words, learner corpus

Abstract

Acquiring multiple senses of a polysemous word perfectly is difficult for language learners. Although some researchers have tried to find the effective way of teaching a polysemous word, few studies have examined how learners develop their knowledge of it. Therefore, this presentation aims to show insights of how Japanese learners of English develop their use of a polysemous word “over”. To achieve this purpose, this study set two research questions;

RQ1: Do Japanese learners of English use “over” the same way as native speakers do?

RQ2: What is the difference between “easy” senses and “difficult” senses?

To answer these two research questions, both quantitative and qualitative research was undertaken. First, a learner corpus, Longman Learners’ Corpus, and a native corpus, British National Corpus, were used to collect sufficient examples of “over” from each proficiency level of the learners and English native speakers. After that, semantic tags of “over” from Tyler and Evans (2001) were used to annotate each example. Then, a correspondence analysis was generated based on the cross-tabulation which organizes the results of the annotation. Finally, the difference between easy senses and difficult senses for Japanese language learners was discussed based on the result of CA and the frequency of each sense.

Correspondence analysis shows clearly that Japanese language learners use “over” differently from English native speakers in terms of its senses. There are some senses which have a strong connection with the learners, such as covering, completion, and repetition senses, while control, focus-of-attention, transfer, preference, the- other-side-of senses have a strong connection with native speakers. In addition, by focusing on the senses which are no occurrence at the beginner level and which are produced little even at the advanced level, this study has succeeded in explaining which senses are difficult for language learners to produce. This study revealed that the protoscene and the senses which are in cluster 4, 5, and 6 in the polysemy network of over by Tyler and Evans (2001) are especially difficult for Japanese language learners. These findings might help teachers when they teach the word “over” to Japanese language learners of English, and help researchers when they assess the learners’ production of English.

【研究発表3】

ムーブと形容詞の振る舞いから見た航空会社プロフィールのディスコース分析

仁科 恭徳（神戸学院大学）

Keywords: ディスコース分析, 形容詞, ムーブ

Abstract

本発表では、タグ付けしたDIYコーパスを用いて量・質の観点から航空会社のプロフィールを分析した結果を報告する。特に、当該プロフィールのムーブ（構造）

と形容詞の振る舞いに焦点を置き、各航空会社が属している3つのアライアンス（スターアライアンス、ワンワールド、スカイチーム）間での量的・質的な違いや、航空会社プロフィール全般に通底している共通性に関して調査した結果を発表する。

ムーブに関しては、ムーブの種類、各ムーブの重要性、典型的なムーブ構造に関して検証する。形容詞の振る舞いに関しては、形容詞＋名詞の連辞的結合 (colligation) や best + AWARD といった優先的意味選択 (semantic preference) に注目して分析を進める。

ビジネス・コミュニティにおいて、プロフィールは代表的なジャンルの一つである。よって、このジャンルにおいて共有され慣習化された言語実態を暴くことで、ビジネス・コミュニティに通底している知識や文化の理解を深めるだけでなく、English for Business Purposes（ビジネスに特化した英語（教育））における教材開発やシラバスデザインの元資料としても有効に活用することができる。

本研究におけるリサーチ・クエスションは以下のとおりである。

1. コーパス・データに基づく各種プロフィールの基本的な異なり
2. 当該プロフィールで使用されているムーブの特定（種類と数）
3. 当該プロフィールにおける各ムーブの必要性の数値化
4. 当該プロフィールにおける典型的なムーブ構造の解明
5. アライアンス間における形容詞の振る舞いの類似性と相違性

1～5のリサーチ・クエスションをもって、当該プロフィールの実態を精緻に暴く。1は当該プロフィールと他のプロフィールとの違いの可視化、2～4は当該プロフィールが担う伝達機能の解明、5はよりマイクロな視点から分析した当該プロフィールの特徴の解明を目的とする。例えば、リサーチ・クエスション2に関しては18種のムーブが認められ、リサーチ・クエスション3に関しては2種のムーブが義務的、3種のムーブが慣習的に使われていたが、発表時には質的な考察も含めて詳細を報告したい。

【研究発表4】

CEFR-J 準拠英語教育用 web コーパスの開発

投野由紀夫（東京外国語大学）
川原田将之（東京工業大学）
渡辺 亮嗣（ネットアドバンス）
星野 守（ネットアドバンス）
奥村 学（東京工業大学）

Keywords: CEFR, web corpus, pedagogical corpus

Abstract

本研究は、ヨーロッパ言語共通参照枠 (CEFR) を日本の英語教育に適用した CEFR-J (投野, 2013) を用いて、英語教育を側面から支援する教育用コーパスを構築するプロジェクトの中間報告である。

手順としては、まず基本分野の小中高生向けの英語サイトを分野別に調査閲覧し、

内容的に適切と思われるサイトの URL リストを作成した。次いで、URL を手がかりに web 上から HTML ファイルを自動取得し、収集した HTML ファイルからヘッダやフッタ等の不要な情報を取り除くため、ブロック単位と BootCat の2方法でテキストを抽出した。センテンス数やタグ情報などの指標をもとに、これらのファイルをスクリーニングした。この時点で、4億5500万語（586,660ファイル）のテキスト・データが取得できた。

次にこれらのテキストを、CEFR-J プロジェクトの資源を用いて東京工業大学奥村研究室で作成した CEFR レベル判定ツール (English Level Checker) を用いて、CEFR レベル (A1~C2) に自動分類した。この判定ツールは、テキスト中の CEFR-J Wordlist に基づく使用語彙レベル、主要文法事項の出現率、などをもとに機械学習されたものである。CEFR レベル判定ツールの評価を、A1~B2各レベル20サンプルを抽出し2名の訓練された評価者により人手で CEFR 評価を行い (Cohen's Kappa=0.72)、機械判定と比較した。結果は A1-A2,A2-B1などの隣接レベルでの誤判定がまだ多かったが、大きく A レベルと B レベルの2クラス判定では70%程度の一致度であった。

最後に小学館コーパスネットワーク (<https://scnweb.japanknowledge.com/>) へ搭載され、新学習指導要領に基づく CAN-DO ベースの英語教育を側面から支援する言語教育資源としての意義、CEFR-J レベルの細分化への対応など、将来的な課題にも言及する。

< Presentation Q&A session 3 >

[研究発表 1]

Issues in Compiling and Exploiting Textbook Corpora

Elen LE FOLL (Osnabrück University)

Keywords: Corpus Design, Coursebooks, Pedagogical Materials, English as a Foreign Language, Classroom English

Abstract

Textbooks are known to be “one of the most important educational inputs” (Pingel 2010: 7) and, as such, have long been cherished objects of research, especially in the social and political sciences. In applied linguistics, page-by-page analysis of textbook language was once a difficult, time-consuming process. However, the development of digital data storage and retrieval enabled Mindt (1987; 1992) to pioneer a new approach to language textbook analysis using computer-readable textbook corpora. Since, multiple studies have highlighted the potential of corpus-based textbook analysis both for materials development and evaluation, and as a means of capturing learner language input and better understanding learner language production (cf. Römer 2004; Meunier & Gouverneur 2009).

However, in practice, compiling textbook corpora can be an arduous task, fraught with potential pitfalls. In this paper, I highlight some of the issues specific to compiling a corpus of textbooks and propose possible solutions. The compiling process begins with the selection of

the materials to be included in the corpus, a process driven by the research questions. I discuss questions pertaining to the design of the sampling frame, corpus size, representativeness, and balance (cf. Biber 1993). Due to the many different font types, colours, and often complex page layouts, automatic OCR (optical character recognition) is usually much more complex for textbooks.

For many corpus linguistic measures, the basic unit is the text. However, textbooks are essentially collections of many different texts. Moreover, textbooks also pose problems due to some texts being extremely short (e.g. instructions), whilst others can span several pages (short story). In addition, what constitutes a textbook must also be defined since educational publishers now offer multimedia packages that include coursebooks, workbooks, teacher's guides, audio and video materials, etc.

These issues in compiling and exploiting textbook corpora are illustrated with examples from the compilation and analysis of a manually annotated textbook corpus of nine series (43 volumes) of secondary school EFL textbooks.

【研究発表 2】

Phrase Frames As an Exploratory Tool for Studying Translation Patterns: A Corpus-Based Descriptive Study

Lukasz GRABOWSKI (University of Opole)

Keywords: phrase frames; phraseology; parallel corpora; English-to-Polish translation; corpus-based study

Abstract

Designed as a proof-of-concept, this descriptive corpus-based study focuses on the construct of phrase frames, originally defined as a contiguous sequence of *n* words identical except for one (Fletcher 2002). Although phrase frames were already used as a means of exploring pattern variability across and within different text types or registers written in English (Römer 2010; Gray & Biber 2013; Fuster-Marquez & Pennock-Speck 2015; Grabowski 2015; Forsyth & Grabowski 2015; Cunningham 2017; Juknevičienė & Grabowski 2018; Lu et al. 2018), there has been no attempt so far to employ that construct as a unit of analysis in descriptive research on translation. More precisely, we aim to verify whether phrase frames found in English source- language texts reveal similar generalizable syntagmatic patterns in Polish translations. Also, we aim to investigate whether the observed translation patterns help highlight lexical and stylistic peculiarities of translations which would otherwise be difficult to capture.

In this study, we use the English-Polish parallel corpus *Paralela* (Pęzik 2016), notably its European Parliament proceedings sub-corpus (Koehn 2005), to identify and describe translation patterns that emerge from one functionally-defined English phrase frames

expressing attitudinal stance (it is * clear that) used as a starting point for the analysis. The findings provided insights into English-to-Polish translation patterns, which revealed that the Polish equivalents are realized with a high degree of regularity and can be generalized into syntagmatic patterns similar to phrase frames. We also obtained valuable cross-linguistic insights into corresponding syntagmatic structures in English and Polish. All in all, the study design is potentially applicable beyond English-Polish translation.

【研究発表 3】

Investigating Publisher Application of Corpus Research on Recent Language Change To ELT Coursebook Development

Niall CURRY (Coventry University)

Robbie LOVE (Aston University)

Olivia GOODMAN (Cambridge University Press)

Keywords: ELT, materials development, corpus linguistics, adverbs, spoken British English, language change

Abstract

Corpus linguistics (CL) has come to occupy an important space in the language teaching industry (see e.g. Römer, 2011), but the specific ways in which English Language Teaching (ELT) publishers use CL research to inform materials development are under-studied, meaning that it is not known whether CL is being used by publishers to its full potential. This study investigates the use of CL research by a major international ELT publisher – Cambridge University Press – by (a) conducting a case study into recent change in adverbs in casual spoken British English; (b) sharing the findings with editors from the publisher; and (c) investigating how the editors use the corpus-informed findings in developing coursebooks. We address two main research questions:

1. How has the use of adverbs changed in casual spoken British English between the 1990s and 2010s?
2. What is the role of the editorial process in developing corpus-informed coursebooks?

Through our analysis of the Spoken BNC1994 (BNC Consortium, 2007) and the Spoken BNC2014 (Love et al., 2017), we find evidence of major recent changes in the usage of frequent adverbs. Overall, adverbs are significantly more frequent in the more recent dataset. Furthermore, there is evidence of noteworthy functional and syntactic shifts in a number of adverbs including ‘like’, ‘so’, ‘just’, ‘well’ and ‘literally’, which, we argue, warrant discussion and presence in the English language teaching classroom.

Following the corpus analysis, we conducted in-depth interviews with the editors and a review of the materials they subsequently produced using the corpus findings (Cowan et al., 2018; Kilbey et al., 2018; Goldstein & Jones, 2019). We evaluate how our findings are

represented in the materials, finding some evidence of effective use of corpora in materials development but revealing limitations in current corpus research approaches which prevent editors from employing CL research more effectively.

【研究発表 4】

Pronominal Ambiguity and Ascriptions of Responsibility in the UK Daily Coronavirus Briefings

Jamie WILLIAMS (Nottingham Trent University)

David WRIGHT (Nottingham Trent University)

Keywords: discourse, corpus-based discourse analysis, political discourse, COVID, coronavirus

Abstract

Within political discourse, pronouns have been highlighted as important linguistic features due to their inherent ambiguity, and their roles in creating distance or closeness and accepting or denying responsibility for actions (Fetzer and Bull, 2008; Mulderrig, 2012). These issues are pertinent to the COVID-19 pandemic, as governments attempt to clearly communicate guidance to the general public, as well as describe steps being taken to slow the virus' spread. Within the context of the United Kingdom, one of the worst affected countries globally, we investigated how pronouns were used by governmental speakers to administer responsibility and whether they contributed to reported criticisms of ambiguity in the government's communications (Oliver, 2020).

A corpus of 92 political speeches, totalling 117,779 words, was constructed based upon official transcripts from the UK government's website. Focussing on the use of the first person plural (1PL) pronouns, 3,045 concordance lines were analysed to identify (1) their referent – particularly if they carried an exclusive (we – the government) or inclusive (we – the country) reading, and (2) the transitivity patterns these pronouns act as Participants in (Halliday and Matthiessen, 2014).

We argue that the UK government uses the inherent ambiguity between the exclusive and inclusive readings of this pronoun to mitigate their own portrayed responsibility for controlling the spread of the virus. We argue that they do so through at least two means. Firstly, when using 1PL pronouns in an exclusive manner, although they represent themselves overwhelmingly as Actors, they obscure the precise details about the measures they are taking. Secondly, when using the 1PL pronouns in an inclusive manner, they represent the British public as co-Actors in processes they have no control over and indeed are usually considered to responsibility of the government themselves.

【研究発表5】

A Corpus Analysis of Prepositional Colligations in Nigerian Legal Discourse

Ibrahim BASHIR (Jubail Industrial College)

Kamariah YUNUS (Universiti Sultan Zainal Abidin)

Keywords: British National Corpus (BNC), corpus linguistics, colligation, prepositional colligations, legal phraseology, Nigerian law corpus (NLC)

Abstract

English is recognised as a national lingua franca of Nigeria. It is adopted as a medium of instruction at almost all levels of the Nigerian educational system, in bureaucracy, the main mass media, national politics, science and technology, corporal companies, two legislative chambers, and the judiciary. The empirical studies using corpus approaches had shown that prepositions are essential linguistic features of legal phraseology. That area was under explored in the Nigerian context. As Nigeria is using English in legal affairs, the present study purports to investigate the representativeness and essentiality of prepositional colligations in Nigerian legal texts. Using a corpus methodology, a new corpus was compiled namely Nigerian Law Corpus (NLC) which contained 546313 word-tokens. The list of prepositional colligations generated from NLC were compared with two reference corpora – British National Corpus of Law (BNCL – 2.2 million word-tokens), and British National Corpus of General Written English (BNCW – 1 million word-tokens) explored using Lextutor (v.8.3). The analysis was aided with the AntConc (v.3.5.7), Lextutor (v.8.3), and Rayson's Log-likelihood Calculator. This paper reports the quantitative parts of a large study. The findings revealed that prepositional colligations were prevalent linguistics features in Nigerian legal discourse. The findings showed that prepositional colligations were under-represented in the NLC compared with BNCL (NLC observed frequency = 3619, relative frequency = 0.66, BNC observed frequency = 17923, relative frequency = 0.81, Over-/under-representation (-), Log likelihood = 134.82). On the other hand, prepositional colligations were over-represented in NLC compared with BNCW (NLC observed frequency = 3619, relative frequency = 0.66, BNCW = 3939, relative frequency = 0.39, Over-/under-representation (+), Log likelihood = 500.43). The study concludes that prepositional colligations were important linguistic elements that formed the phraseological profiles of Nigerian legal discourse due to their prevalent distributions and multifarious semantic and pragmatic functions.

【研究発表6】

**Factuality and Conditional Sentences with Indicative Mood in Subordinate Clauses:
A Corpus-Based Study**

Barrios LEYRE (Universitat de Lleida)
Vázquez GLÒRIA (Universitat de Lleida)

Keywords: factuality, if-conditionals, Indicative, Spanish, annotation.

Abstract

Factuality (or certainty) is the notion that refers to the speaker's commitment to the truthfulness of a situation, which is expressed through different linguistic markers. This study belongs to TAGFACT project whose aim is to analyze and automate the interpretation of factuality in Spanish in journalistic texts.

A conditional sentence consists of two clauses, a main clause (apodosis) and a subordinate clause (protasis), which are linked by a conjunction typically used for hypothetical situations (Quirk et al., 1985; Sweetser 1990, Dancygier 1993; Montolío, 1999; RAE 2009). One of our aims is to analyze the factuality of both clauses of some conditional sentences in Spanish, particularly those conditionals with protasis with Indicative and introduced by the conjunction *si*, which in English is *if*. Also, we pretend to propose linguistic clues for the automatic annotation of factuality for both clauses.

The protasis of a conditional sentence typically do not describe factual situations, but this is just in some types of conditionals. In addition, the factuality of apodosis has hardly been studied, which is very important, as we will see later.

For Spanish, there are studies that analyze the interpretation of the protasis of these sentences in terms of factuality, but not of the apodosis. Different bipartite (Veiga and Mosteiro, 2006) or tripartite (Montolío 1999) factual classifications have been made. Our study is based on the latter, since it goes deeper into formal issues and distinguishes between real, potential and unreal conditionals from the point of view of the protasis.

we have analyzed 241 if-conditionals from 3 different corpus: the corpus Now (Mark Davies, 2019), CORPES XXI (Real Academia de la lengua Española, 2020) and CREA (Real Academia de la lengua Española, 2015) whose protasis include indicative tenses: presente de Indicativo, pretérito imperfecto de Indicativo, pretérito perfecto simple and pretérito perfecto compuesto. In total, 25 temporal combinations have been analyzed.

Firstly, the interpretation of the factuality of the protasis has been revised, extending in some cases the current bibliographical proposal, and secondly, a factual interpretation has been proposed for the apodosis.

The result of this study is the proposal of a series of linguistic rules to formalize the interpretation of the factuality of the two clauses of the conditional sentences that present protasis with indicative in Spanish, thus completing the previous study done on the subjunctive (Barrios and Vázquez, 2020).

【研究発表7】

**Corpus Linguistics Tools for the Creation of Linguistic Resources
That Support the Internationalisation of Tertiary Education**

Rosana VILLARES (University of Zaragoza)

Keywords: corpus linguistics, teaching guides, glossary, translation, internationalisation**Abstract**

In the last decades, universities have developed different internationalisation strategies based on the English language, such as English-medium instruction courses or the translation of institutional websites, with the purpose of improving their international visibility and attracting international students (Dearden, 2014; Ferguson, 2007). However, the literature (e.g. Ferguson et al., 2011; Pérez-Llantada et al. 2011) has consistently reported the multiple difficulties faced by non-English native scholars when using English for their academic tasks. As part of the internationalisation strategy of a medium-sized Spanish university, lecturers must translate to English their modules' teaching guides. Given this situation, this study proposes the use of corpus linguistics tools to create a series of support resources that facilitate the translation of those documents. A corpus of 113 institutional reports written by the university and verified by the Spanish National Agency for Quality and Accreditation. These documents were gathered because they included all the information that later on the lecturers have to use in their teaching guides of both undergraduate and graduate degrees. In this way it was possible to familiarise with the official terminology that lecturers should use. The AntConc v3.5.7 software was used to extract frequency lists and concordances to identify the lecturers' recurrent terminology and phraseology. Based on the quantitative results, three resources were created. Firstly, a glossary of the teaching activities with their translation and definition so that lecturers could use them accurately. Secondly, a glossary with the most frequently used expressions written in Spanish, their equivalent in English, and a concordance line that illustrates the use of that specific term in context. Lastly, several teaching guide templates were created to homogenise the teaching guide genre while simultaneously allowing variation according to disciplinary differences. By sharing these resources with the lecturers, it was expected that some of the translation challenges they may face (terminology inconsistencies, literal translation, target readership) (Hurtado, 2001; Parra Galiano, 2005) would be overcome.

【研究発表 8】

Speaking of Extinction: A Comparative Corpus-assisted Analysis of the Environmental Framing in Climate Fictions and the News on the Net

Hakan CANGIR (Ankara University)

Taner CAN (TED University)

Keywords: Cli-fi, Semantic Domains, Collocations, Framing, Metaphors**Abstract**

The cognitive linguist George Lakoff developed an alternative view of metaphor which has long been relegated to the context of creative writing. In his view, metaphors are part of cognitive frames we use in our everyday lives to understand abstract concepts in terms of concrete objects and activities. Straddling contemporary cognitive theories and corpus linguistics, the present study seeks to compare the (metaphorical) frames in the climate fictions (aka cli-fi) and the news on the Internet by identifying the key semantic domains, collocations and key words in context. For this comparative analysis, we compiled a corpus of 26 novels by 26 novelists and exploited the Climate Change in the News Corpus provided by Lancsbox. We used the tools Wmatrix for semantic tagging and Lancsbox and Voyant tools to report collocational patterns and key words in context. The theoretical and methodological approach of the study was mainly based on Lakoff's framing theory and ecocriticism. As the initial stage of the research, we designated two key semantic domains, namely "Green Issues" and "Universe," and explored semantic collocations and key words in context in the two corpora. When the two corpora are compared, the preliminary results indicate that cli-fi writers employ a more powerful language (e.g. "ecological terror") to bring up a new perspective on the environmental crisis we are going through. These writers' darker and emotionally disturbing tone is in stark contrast to the neutral, disinterested tone we find in the news corpus. News writers seem to use a more business-oriented language where we find the words with strong political and economic connotations. Subsequently, the preliminary results drawn from the research is partly in line with Lakoff's concerns about the cognitive frames we create about environmental problems, and his emphasis on the role of storytelling in changing these false frames.

【研究発表 9】

Exploring the Use of Hedges in Academic Writing: A Corpus-Based Analysis Between Chinese TESOL Students and Expert Writers

Yagang CHEN (The University of Edinburgh)

Keywords: hedges, academic writing, Chinese TESOL postgraduates, published research articles

Abstract

The ability to write academic papers in English has been acknowledged as an important criteria to evaluate academic profession around the world (Tang, 2012). In academic writing, appropriate use of hedges is critical. However, the ability to express doubt and certainty appropriately in English is generally acknowledged as a difficult task for language learners. The present study investigates how Chinese TESOL postgraduates employ lexical hedges in their academic writing in comparison to professional writers, attempting to help improve their writing skills. Using Hyland (1998a)'s taxonomy, a quantitative corpus-based study was conducted to examine hedges based on grammatical class and function. The findings demonstrate that : 1) Chinese TESOL postgraduates use hedges extensively in their academic writing. 2) Chinese TESOL students display similarities in overall frequency of hedges and choice of some hedge items. 3) Both groups rely on content- oriented hedges more than reader-oriented hedges. Differences are also detected. Chinese TESOL students employ significantly more hedges than expert writers, particularly with modal verbs and adverbial hedges.

Different preferences of individual hedge items by the two groups are also presented. Pedagogical implications generated from the main findings are concerned with material design and teaching instructions for EAP and academic writing courses at tertiary level.

【研究発表10】

A Comparative Study of Transition Markers in Thesis Abstracts From Chinese Undergraduates and RA Abstracts From Prestigious Journals

Xiaojing CHEN (The University of Edinburgh)

Keywords: genre analysis, metadiscourse, transition markers, thesis abstracts, research article abstracts

Abstract

The researcher conducts a comparative study between thirty thesis abstracts written by Chinese undergraduates and thirty research article abstracts collected from top-ranked journals in applied linguistics. Abstracts become an independent genre which can reflect language proficiency and critical thinking of the researchers. This study focus on the frequency of using transitions (e.g. “but” and “therefore”), which are the metadiscourse markers particularly used to structure the text by explicit connection between the main clauses. Five types of transition markers are identified and log-likelihood scores are used to investigate whether there are significant differences between the two kinds of abstracts. Subsequently, two interviews are also conducted to explore the in-depth reasons for the LL scores. The overall frequency of using transitions is not significantly different. However, according to the interviews, Chinese undergraduates sometimes overuse the transitions because it is easier to structure their articles which reveals their abundant vocabulary but deficiencies in text organization. The low

frequency in the use of comparison may be attributed to the lack of questioning competence in examination-oriented education. Based on the results, the researcher believes that the focus for Chinese EFL learners in academic writing should be shifted from grammatical knowledge to research logic and critical thinking.

【研究発表11】

Shared Experience: From I-Illness To We-Illness? Narrative Informed Corpus Linguistic Analysis of a Moderated Online Mood Disorders Forum

Mat RAWSTHORNE (The University of Nottingham)

Keywords: peer support, corpus linguistics, narrative

Abstract

Analysis of interactions in a supervised online mutual aid platform serving people with stress and low mood. Key questions: Are the forum participants sharing their stories, a common strategy for soliciting social support in online communities and does this sharing process assist in recovery by destabilising their narratives?

Methods: A combination of computer-assisted corpus linguistic analyses were employed to examine the structure and nuances of the exchanges, attempting a ‘big qual’ study of broad social processes while also attending to important detail of word use. Results: Although people turn to peer support to help and be helped through sharing stories, there does not seem to be much surveillance for the innovative moments that would shift from being stuck in the ‘Same Old Story’. While platforms like Big White Wall can provide a pause in experience that facilitates separation between the person and the problem, there is no evidence of them stepping back to frame new perspectives to make room for counter narrative. Pennebaker et al (2003) proposed that healing could be detected by shifts from singular to plural personal pronouns, and verb tense changes (Tausczik & Pennebaker, 2010). Closer scrutiny of the interactions points to people talking at each other rather than to or with each other.

【研究発表12】

As Good as the Men? A Corpus-Based Discourse Analysis of Media Representations of Athletes Competing in the New Women’s Australian Football League

Melissa KEMBLE (The University of Sydney)

Keywords: gender bias, sports news, discourse analysis, Australian Rules Football (AFL), evaluation, corpus linguistics

Abstract

This study investigates how players in the newly established women's Australian Football League (AFLW) are represented in the print news media, with respect to patriarchal discourses. Australian Football not only has one of the largest sporting communities across the country, but also boasts the fourth highest live crowd attendance globally (Gullen, 2015). As with many other physical contact sports, the professional league originated exclusively as a male sport, until the new women's league officially commenced in 2017. This research is based on a 87,500 word corpus of AFLW news articles published in the Herald Sun newspaper, one of the most widely-read papers across Australia. The data are taken for the 12-month period commencing June 2016, excluding articles published during the nine-week season. A reference corpus is also used.

Combining corpus linguistics with text analysis of appraisal (i.e. evaluative language), I analyse the AFLW corpus to identify whether there is evidence of patriarchal discourses (i.e. othering, objectification, trivialisation, stereotyping). I also compare the media coverage before the season to the coverage after the season in order to identify whether there are any significant differences in the way players are represented. The keyword analysis reveals a potentially positive shift away from previously documented patriarchal discourses; however, the text analysis reveals a disproportionate focus on gender stereotypes and othering. The (female) athletes are frequently negatively appraised in direct comparison to their male counterparts, which thus positions the professional Australian Football sporting domain as preferably male. This research contributes to existing linguistic research on gender bias by exploring how female athletes entering a male-dominated 'masculine' sport (Koivula, 2001) are portrayed in the media, with respect to patriarchal discourses. It provides a foundation for future linguistic research into representations of athletes in sports reporting using a mixed-method approach combining corpus-linguistics with CDA.

October 4 (Sun), 2020

< Plenary Talk 2 >

Plenary Talk 2: NLP Beyond NLPers – The Many Faces of NLP in Academia and Real-World

Sowmya VAJJALA

Data Scientist and NLP Research Officer (National Research Council Canada)

Biography

Sowmya Vajjala currently works as a researcher in Digital Technologies at National Research Council, Canada's largest federal research and development organization. She has worked in the area of Natural Language Processing (NLP) over the past decade in various roles – as a software developer, researcher, educator, and a senior data scientist. Her research interests lie in multilingual computing and educational applications of NLP. She recently co-

authored a book: “Practical Natural Language Processing: A Comprehensive Guide to Building Real World NLP Systems”, published by O’Reilly Media (June, 2020). She is interested in learning more about the relevance of NLP beyond research both in industry practice as well as in other disciplines, through inter-disciplinary research.

Abstract

Natural Language Processing is an active area of research and its impact is also seen in many day-to-day applications we use, from generic tools such as email software to specialized ones such as language learning apps. Apart from being an active area of enquiry in itself, NLP methods are widely used in many disciplines, from linguistics to economics, from psychology to plant science. In this talk, I will introduce some common NLP practices, show how NLP differs between academia and industry, and discuss areas where NLP is useful beyond its home turf, including its use in corpus linguistics research. Drawing on my experiences as an NLP researcher and instructor, I will also touch upon what NLP can learn from corpus linguistics and other areas of study, and what is needed to train diverse groups interested in using NLP methods in their work.

< Presentation Q&A session 4 >

【研究発表1】

A Corpus-Based Analysis of Ongoing Change in the Adjective Amplifier Systems of Hong Kong, Indian, and Philippine English

Martin SCHWEINBERGER (The University of Queensland)

Keywords: Adjective Amplification, Hong Kong English Philippine English, Indian English, Conditional Inference Trees, Variationist Sociolinguistics, Language Change

Abstract

This study focuses on one of ongoing change in adjective amplifiers (very, really, so, etc.) in Hong Kong (HKE), Indian (IndE), and Philippine English (PhiE) based on data from the International Corpus of English.

While previous research on changes in amplifier systems has successfully applied multivariate methods and unearthed intricate interdependencies and highly systematic trajectories of change in inner circle varieties of English (e.g. D’Arcy 2015; Tagliamonte & Denis 2014), only few studies (e.g. Fuchs & Gut 2016) have analysed ongoing change in adjective amplification in Asian varieties of English. The current study adds to existing research in focusing on change in adjective amplification in HKE, IndE, and PhiE.

One of the most consistent findings in previous research on changes in adjective amplification in informal spoken discourse has been the replacement of very by really. The present study uses Conditional Inference Trees to assess if this trend holds true for HKE, IndE,

and PhiE and if the underlying factors that drive this change in inner circle varieties are also at work in these Asian English varieties.

The analysis shows that the amplifier systems of HKE and IndE are very stable and amplifier choice in these varieties is determined predominately by intra-linguistics factors (adjective type, syntactic context, semantic category). In contrast, the amplifier system of PhiE shows notable signs of ongoing change which is driven predominately by social factors (age and gender of speakers).

The results indicate that during stasis and initial stages of change, language-internal factors determine amplifier choice while social factors become more important once changes have reached mid-range. The paper argues that once certain variants (really and so) gain social meaning, it is this social meaning (the association with specific social groups) drives and accelerates change as speakers want to associate with social groups that carry covert prestige.

【研究発表2】

Becoming a President: A Diachronic Study on the Language of Brazilian President Jair Bolsonaro

Marine Laísa MATTE (Univates University)

Larissa GOULART (Northern Arizona University)

Simone SARMENTO (Federal University of Rio Grande do Sul)

Rozane Rodrigues REBECHI (Federal University of Rio Grande do Sul)

Keywords: Discourse analysis, political speech, diachronic variation.

Abstract

Since the election of far-right candidate Jair Bolsonaro as president of Brazil, several newspapers have written about the president's discourse and his use of language (cf. Beirão, 2019; Lago, 2018). Some claim Bolsonaro speaks the "language of the people" or "the language of social media", among other coinages. Nevertheless, none of these reports have attempted to conduct a systematic analysis of Bolsonaro's speech. Following corpus-assisted discourse analysis studies (Baker, 2006), in this presentation, we conduct a diachronic investigation of Bolsonaro's political speeches during his seven terms in the House of Representatives, which started in 1991. Our goal was to identify the themes that became popular as his political career progressed.

The corpus was compiled from the official records of the government and divided into seven subcorpora representing each mandate. Multi-word units (MWU) were extracted using Sketch Engine. Then, key-MWUs were identified for each of the seven mandates using the speeches of other representatives as a reference corpus. Overall, 1,641 MWU composed of 2 to 4 words were identified as key, with 279 of those occurring in all seven subcorpora. These MWUs were then classified according to semantic categories that emerged from the data, such

as legal matters, armed forces, money and natives' rights. Preliminary results show that over time references to money and communism have declined, while references to education, punishment, and to electoral matters have steadily increased. The results of this quantitative and qualitative study can help us better understand the construction of the far-right discourse in Brazilian Portuguese as Bolsonaro is a representation of right-wing politics, and encourage future comparisons across languages.

【研究発表3】

Media Representations of 'Leftover Women' in China: A Corpus-assisted Critical Discourse Analysis

Yating YU (Department of English, The Hong Kong Polytechnic University)

Keywords: Leftover Women; Media Representations; Corpus-assisted Critical Discourse Analysis; Gender Ideologies; Chinese English-language News Media

Abstract

The term 'leftover women', commonly referring to single women older than 27, has been in popular use in Chinese media since 2007. This study investigates how leftover women are linguistically represented in the English-language news media in China by employing a corpus-assisted approach to critical discourse analysis (CDA). A specialised corpus of 303 English news articles (i.e., 236,254 words), covering the years between 2007 and 2017, was built for this purpose. This study adopts a three-step procedure (i.e. identification, interpretation, and explanation) to examine the immediate co-texts of the lemma LEFTOVER WOMAN by combining the corpus linguistics (CL) concept (i.e. Sinclair's Meaning Shift Units) and techniques (i.e. collocates and concordances) and CDA approaches (van Leeuwen's sociosemantic approach, Charteris-Black's critical metaphor analysis, and Lazar's feminist critical discourse analysis). These findings shed light on media representations of leftover women, the contested ideologies emerging from these representations, and how shifting gender politics and identity shapes and is shaped by media in the world's most populous nation. Additionally, this study contributes to the growing literature of corpus-assisted CDA in the domain of gender representations by proposing an analytical framework that is potentially applicable to future research.

【研究発表 4】

A Novel Approach to the ESP Keyword List: 2800 Entries with Frequent Lexical Bundles for Data-Driven Learning

Michael HENSHAW (Hokkaido University)

Keywords: AntConc, DDL, lexical bundles, n-grams, ESP**Abstract**

Here I present and describe the creation of a wordlist designed to go in the hands of students. Many English for Specific Purposes (ESP) wordlists are constructed with coverage metrics and efficiency in mind, often for use by instructors or materials developers, meaning these lists and associated corpora are scarcely seen by students. Furthermore, while general academic multiword lists exist, those designed for specific fields are lacking. Thus, as a model to fill these gaps, I present the Keywords of One Health Biomedical Sciences (KOBS), a 2800-lemma database derived from the One Health English Corpus, a 2.8 million-token corpus of 651 research articles. Both materials were created for and distributed to L2 first-year PhD students of veterinary medicine, a multidisciplinary field guided by the One Health approach whose interests overlap with medicine, ecology, and other health sciences. To my knowledge, this is the first large vocabulary list or corpus developed for veterinary medicine. Additionally, from inception, KOBS was designed for the student-as-user; it is an easily navigable 2 MB Excel file with over 10 tabs of curated sublists. Each entry includes: most-used lexical bundles both left and right of the node, thematic category (e.g. lab technique, comparison, biochemistry), highest correlated subcorpus (e.g. Introduction, Infectious Diseases), and more. KOBS is a data-driven learning tool for writing research papers whereby students may, for instance, filter the Discussion sublist for category ‘transition’ to return 31 words disproportionately used in the discussion section: #1) also. freq./text=3.38: “it has also been shown to”, “not only in X, but also in Y”; #31) unfortunately. freq./text=0.04: “Unfortunately, this (method does not)”. In short, this enhanced keyword list offers a method for writers to bypass direct interaction with corpora and achieve quick solutions.

【研究発表 5】

Contrasting Narratives: The Greek Financial Crisis in Newspaper Editorials

Will LINGLE (University of Aizu)

Keywords: Critical Discourse Analysis, Corpus Assisted Discourse Studies, News Discourse**Abstract**

Critical discourse analysis (CDA) provides approaches to explicitly political analyses of

texts in their social contexts, though CDA has attracted criticism for overinterpretation of results from analyses of small numbers of texts (Widdowson, 2004). Central to the critical interest in mass media representations is the claim that media texts exert ideological effects on readers in cumulative fashion through repetition of similar characterizations of events and social actors. Corpora offer means of investigating CDA claims, which may be combined in an approach known broadly as corpus-assisted discourse studies (CADS) (Baker, 2006; Partington, 2003). This study began with a Systemic Functional Linguistics (SFL) transitivity analysis (Halliday and Matthiessen 2014) of two US newspaper editorials on the Greek financial crisis, which revealed differences in how the two texts expressed sympathy for key social actors and attributed blame for the crisis. To test whether the portrayals in these two texts represented consistent patterns in the editorial pages, corpora of over half a million words of editorial texts from each paper from 2013-2015 were compiled. Texts focusing on the Greek crisis were retrieved from these using AntConc (Anthony 2011) to produce smaller specialized corpora. Using these, the transitivity analysis was scaled up to focus on all clauses representing one of the three main groups of social actors in the crisis. In each corpus, dominant patterns of representation that could be called ‘narratives’ or schema emerged which revealed that the two newspapers expressed sympathy and attributed blame for the crisis in ways consistent with the two sample texts. The New York Times expressed sympathy for the Greek people and blamed the EU creditors, while the Washington Post blamed the Greek government and praised the EU creditors. Corpora proved valuable informants into broader patterns of discourse, and as a means of reducing researcher bias.

【研究発表6】

The Discursive News Values of the 2017 Marawi City Crisis: A Corpus-Assisted Multimodal Discourse Analysis of Selected Newspaper Reports

Wilfred Gabriel A. GAPAS (University of Santo Tomas)
Rachelle BALLESTEROS-LINTAO (University of Santo Tomas)

Keywords: Marawi City, terrorism, news values, corpus-assisted multimodal discourse analysis (CAMDA), press photography

Abstract

The reporting of both terrorism and conflicts is extensively investigated in past media discourse studies across contexts. However, a dearth in the literature exists regarding newsworthiness and its establishment in various forms of news discourse. Based on this niche, this thesis sought to closely examine how the 2017 Marawi City siege is constructed in print news reports as newsworthy. Specifically, it was aimed at determining (a) the linguistic and visual resources used to construe newsworthiness, (b) the news values used to construe the event’s newsworthiness, and (c) the existing multimodal relationship between language and images. Sixty news reports from three Philippine broadsheets were analyzed using Bednarek

and Caple's (2017) discursive news values framework. This multimodal corpus was examined with multiple methods, namely: (a) collocation analysis of MARAWI and its top three first order collocates, (b) manual analysis of verbal and visual devices utilized for each news value, and (c) the verbal and visual relationship in terms of each discursive news value. A scrutiny of the corpus reveals how local broadsheet journalists constructed the crisis as newsworthy. Firstly, MARAWI is associated closely with collocates that establish Proximity, Negativity, Eliteness, Personalization, Timeliness, Positivity, and Consonance; through collocates city, said, and Maute, recurring second order collocates of Negativity and Eliteness were largely identified. Regarding the manual analysis of texts and images, a range of verbal and visual devices were identified to construct news values except for Aesthetic Appeal. Secondly, an intrasemiotic assessment of news value trends in both verbal and visual data reveal that while the former constructs a concrete set of news values (i.e., Proximity, Timeliness, Superlativeness, and Eliteness), images appear to lack this feature. Lastly, when assessed for their multimodal relationship, texts and images appear to exhibit harmony, where texts and images complement news values with each other. How the crisis' newsworthiness is constructed in the corpus is believed to have implications on news reporting principles and the role of the media in shaping public awareness of events.

【研究発表7】

Comparing Measures of Directness in Corpora of Essays Written by Iraqi EFL Learners, native English-Speakers, and Advanced ESL College Students

Ashleigh COX (Georgia State University)
Eric FRIGINAL (Georgia State University)
Sabah S. MUSTAFA (University of Baghdad)

Keywords: learner corpus, comparative corpora, EFL writing

Abstract

Learner corpora offer valuable insight on variation in EFL writing. This study compares markers of directness and indirectness in corpora of argumentative essays written by Iraqi students learning English, native English-speaking undergraduate students in the United States, and advanced nonnative English-speaking undergraduate students in the United States. The advanced nonnative English-speaking group is included as a successful nonnative model for comparison. There are many previous comparative corpus studies, but Iraqi EFL writing remains under-researched. Ortega (2017) argues that there is a need to include more under-researched populations of language learners in SLA studies, and a learner-corpus approach is one way to address this gap by expanding our knowledge of inter-language writing patterns (Gilquin & Granger, 2015). One area of writing that can be studied using learner corpora is cultural variation in argument styles.

Comparing learner writing with successful nonnative English writing can also help

inform teachers and researchers about how to help English learners advance. This can be particularly useful for research on helping students with TOEFL essays.

In this study, corpora of argumentative essays written by 100 Iraqi students learning EFL, 100 native English-speaking undergraduate students in the United States, and 100 advanced ESL undergraduate students in the United States were compared. Each student participating wrote two essays, one on the prompt “Write an essay based on this statement: Nowadays, people put too much emphasis on personal appearance and fashion” and another on the prompt “Write an essay based on this statement: The best way to ensure a good future for yourself is to plan carefully while you are still young.” The essays were scored using TOEFL rubrics, and linguistic features were tagged using the Multidimensional Analysis Tagger (Nini, 2014), which use a tagging system modeled after the Biber Tagger. The use of hedges, possibility modals, downtoners, emphatics, and amplifiers were compared for essays written by each group of students, taking into consideration the effect of the prompt. The three groups differed significantly in their use of hedges, possibility modals, downtoners, and emphatics, and they differed moderately significantly in their use of amplifiers. Prompt had a significant impact on the use of hedges and downtoners. Implications for teaching EFL writing, with a special focus on learners from Iraq, are discussed.

【研究発表 8】

Sky High: Building a Corpus of English for Flight Training

Andrew SCHNEIDER (Embry-Riddle Aeronautical University)

Rachelle UDELL (Georgia State University)

Eric FRIGINAL (Georgia State University)

Keywords: Aviation English, English for Specific Purposes, ESP, Corpus development, flight training, policy

Abstract

Corpus research in the field of Aviation English has largely focused on communication between professional pilots and air traffic controllers (see Bieswanger, 2016; Borowska, 2017). However, there has yet to be a thorough corpus analysis of discourse between student pilots and their flight instructors. Furthermore, it is currently unknown to what extent the high-stress, high-stakes environment of in-air flight training influences multilingual student pilots’ L2 development or if factors such as stress and time pressure impact learners’ ability to successfully communicate in this specialized register of English with both familiar and unfamiliar interlocutors. Guided by Biber and Conrad’s (2009) framework, a situational analysis of flight training discourse was conducted at a prominent aviation school in the United States. This analysis revealed that flight training operations involve oral, simulated flight, and in-air flight activities featuring a diverse range of communicative interactions between student pilots, instructors, school support staff, and local air traffic control. The resulting Corpus of

Flight Training (CFT) is the only representative corpus of authentic discourse used in flight training to date. It is comprised of audio and video recordings of one-on-one, instructional communication between ESL student pilots from a variety of L1 backgrounds paired with both mono- and multilingual English-speaking flight instructors. The CFT currently guides the development of pre-training assessments and language support programs for accepted students (classroom instruction, one-on-one tutoring, etc.) and will potentially inform macro language policy decisions for all FAA-certified flight training programs in the United States. The unique nature of this context has presented an equally distinctive set of challenges in the development of the CFT. Nevertheless, collecting corpus data in high-stress environments is both important and beneficial. This presentation explores these challenges and benefits and details the necessary foundations of policy and procedure needed for data collection in high-stakes occupational training scenarios.

【研究発表 9】

The Construction of Coronavirus in English-Language Indonesian Newspapers: A Corpus-Assisted Discourse Analysis

Muchamad Sholakhuddin AL FAJRI (Universitas Airlangga)

Keywords: Media representation, Coronavirus, Covid-19, discourse analysis, corpus linguistics

Abstract

This study aims to examine the construction of Covid-19 in English-language Indonesian newspapers by using corpus-assisted discourse analysis. The data were collated from news articles containing the word “Coronavirus” or “Covid-19” published in two English-language Indonesian news media from January to July 2020. This resulted in 3055 texts with 1,233,822 words. Collocation analysis was employed to reveal discourses by focusing on verbal and adjectival collocates of “Coronavirus” and “Covid-19”. A span of five words on either side of the search term was set and MI score was used as a measure to calculate the collocates. A collocational pair that occurs less than 10 times and has an MI score less than 3 was excluded. The top 100 resulted collocation lists were then grouped thematically on the basis of their semantic meaning, adopting UCREL Semantic Analysis System (USAS) (Piao et al., 2015). Concordance analysis was also carried out to better interpret the collocates and to examine them more qualitatively. My preliminary findings suggest that Covid-19 is constructed as an enemy in war/conflict by using WAR metaphors (e.g. fight, combat, and battle). Portraying this pandemic as a war may force people to be obedient to the government Covid-19 policies as evidence of their patriotism. However, this may also lead to an authoritarian government. For example, the Indonesian government imposes Regulation in Lieu of Law (Perppu) No. 1/2020 on the response to the COVID-19 pandemic that includes the impunity (both civil and criminal law) for policy-making officials in taking any extraordinary measures, which can be

misused by the government. Coronavirus is also represented as a deadly disease (example of collocations: kill, threaten, deadly, contagious, and fast-spreading). This portrayal may help make people more aware of the danger of Covid-19. However, it also can create unnecessary fear and stigma. For instance, some people reject the burial of Covid-19 dead in the cemetery near to their residence, in spite of using Covid-19 burial protocols, because they are afraid of being infected.

【研究発表10】

NICT JLE コーパスを用いた日本人英語学習者のイラスト描写における 習得レベル別特徴分析

寺田 里紗（東京外国語大学大学院生）

Keywords: 学習者コーパス, スピーキング評価, 発話分析, 話し言葉コーパス

Abstract

近年、学習指導要領では英語力の育成に「思考力・判断力・表現力」が強調され、単に知識・技能を身につけるだけでなくそれを場面・状況に合わせてどのように使うか、に関して注目が集まっている。本研究ではこのような観点から、日本人英語学習者のスピーキングにおけるイラスト描写タスクに関して、英語力レベルの異なる学習者の発話を、レベル別の言語的・談話構成的特徴に焦点を置いて分析した。

NICT JLE コーパス（和泉他 2004）のイラスト描写タスク中のレストラン・シーンが描かれているイラストを分析対象として選択した。まず、ほとんどの学習者が言及する中心的場面に焦点を置いて、その場面に含まれている主要な描写要素（人物・服装・位置・行為・表情・背景など）のリスト化を行った。次にこれらの関連部分の描写に関して、SST レベル別学習者集団の産出データを分析し、総語数、文数などの全体量、個別の描写要素の言及割合および発話分量、使用言語表現などについて観察・分析を行った。また基準として母語話者データも同様の分析をし、比較を行った。

結果として、学習段階が上がるにつれて、描写量、描写要素の割合、使用言語表現の多様性、構造の複雑化などが増加傾向になることが明らかになった。発表では、それらの詳しい特徴と SST レベルとの対応関係を質的分析も含めて報告する予定である。

< Presentation Q&A session 5 >

【研究発表 1】

A Corpus Stylistic Comparison of Speech Presentation of 21st-Century Present-Tense Fiction and 20th-Century Past-Tense Fiction

Reiko IKEO (Senshu University)
Masayuki NAKAO (Tottori University)
Eri SHIGEMATSU (Hiroshima University)

Keywords: corpus stylistics, direct speech, past-tense fiction, present-tense fiction, speech presentation

Abstract

In contemporary fiction, the use of the present tense for narration is becoming more common; almost 30% of the titles shortlisted for the Man Booker prize in this century are written in the present tense instead of the past tense. This paper compares direct speech presentation in present-tense fiction of the 21st-century with that in 20th-century past-tense fiction and identifies its stylistic features. This is part of our research project which compares speech, writing and thought presentation in these two textual divisions. The data of this research consists of two corpora. One is a corpus of present-tense narratives (the PREST corpus) and the other is of past-tense narratives (the PAST corpus). The total word count of the PREST corpus is 87,901 while the PAST is 88,319. Both the corpora were annotated using two annotation systems: an online automatic part-of-speech tagging device Sketch Engine (Kilgarriff et al. 2014) and an XML-conformant mark-up system, categorizing speech, writing and thought presentation, which derives from the model developed by Semino and Short (2004).

Speech presentation in the PREST and the PAST corpora respectively accounts for 45.6% and 48.2% of all the tags in each corpus, which suggests that almost half of the texts depict characters' speech. In both corpora, the direct forms account for approximately 29% of all the tags while indirect forms approximately 6%.

However, an examination of the sub-corpus of direct speech in each corpus reveals that speech presentation in present-tense narrative and that in past-tense narrative are qualitatively different. In direct speech in present-tense narrative, present-tense verbs are overused while past-tense verbs are underused. The frequent use of present progressives in characters' direct speech seems to reflect their concerns about the immediate present. In addition, characters' speech styles seem to be more interactive, more responsive to their addressees.

【研究発表 2】**Performance Evaluation of Automated CEFR Level Classification Tools**

Wei-Tung WANG (Tokyo University of Foreign Studies)

Yukio TONO (Tokyo University of Foreign Studies)

Keywords: tools evaluation, automated CEFR level scoring, spoken language assessment**Abstract**

There is a growing interest in using the Common European Framework of Reference for Languages (CEFR) for classifying teaching points (vocabulary, grammar, and texts) into different proficiency levels. Recently, several automated tools are available for judging the CEFR levels based on text characteristics (e.g. Text Inspector, CEFR-based Vocabulary Level Analyzer, among others). The present study aims to critically evaluate the performance of different automated tools for CEFR-level classification of texts.

To this end, the TUFUS-ELC Spoken Corpus was used, which is a collection of oral interview transcripts with CEFR gradings judged by two professionally trained raters. According to the procedure of the Cambridge English Assessment, when two raters have different ratings, a third rater will be consulted for further evaluation to reach an agreement. We basically followed the same procedure. This spoken corpus can serve as a gold standard for comparing different tools' performance. For the performance evaluation, 184 samples were selected from the major CEFR levels, A2 to B2, in order to be processed by three classification tools: (1) Text Inspector, (2) CEFR-based Vocabulary Level Analyzer (CVLA, Uchida & Negishi, 2018), and (3) English Level Checker (Hayashi, et al. 2017). The results were compared against the human raters' judgements as the gold standard.

The results show that the overall CEFR level classifications turned out fairly accurate across the tools, although each tool showed strengths and weaknesses in dealing with different aspects of text characteristics such as how to deal with texts with different length, the proportion of CEFR level vocabulary, the robustness against off-list words, among others. Also, qualitative analyses will be made in order to determine what textual properties affect the inaccurate estimation of CEFR levels compared to human ratings. Possible implications and future directions of automated CEFR-level text classification will be discussed.

【研究発表 3】**Extending Vocabulary Profiling to Languages Other than English**

Laurence ANTHONY (Waseda University)

Natalie FINLAYSON, Emma MARSDEN, Rachel HAWKES, and Nick AVERY
(National Centre of Excellence for Language Pedagogy, University of York)

Keywords: NCELP, vocabulary, profiling, non-English, tools

Abstract

Vocabulary profiles of corpora are often created as a step towards creating and/or modifying pedagogic materials for a target learner audience. Two of the most used desktop vocabulary profiling tools are Range and its more modern equivalent AntWordProfiler. For online vocabulary profiling, Web VP tool, which is part of Compleat Lexical Tutor is a popular alternative. All these tools can in theory be used to profile texts of any language. However, they rely on levelled vocabulary lists where each item in the list is grouped according to its “word family”, “flemma”, or lemma category. They also rely on each item in a list being a single string of characters (i.e words). These limitations introduce problems when attempting to profile languages such as English and French (and almost all other languages) which are composed of both single- and multi-word units. They also hugely complicate the process of vocabulary profiling for languages with a high degree of declension such as German and Spanish. In this presentation, we will first discuss the problems of vocabulary profiling in English and languages other than English. Next, we will explain how an existing desktop profiling tool was adapted for use at the National Centre for Excellence for Language Pedagogy (NCELP), UK to assist researchers in the creation of curricula specifications for the teaching of Spanish, French, and German vocabulary and also teachers hoping to implement these specifications. Then, we will explain the next stage of the project, which is to develop an open-access, online version of the tool.

【研究発表4】

現代英語における he/she of NP の振る舞いー COCA での調査からー

松田 佑治 (立命館大学)

Keywords: 人称代名詞 (主格), of 句による後置修飾, he/she of NP, 同格用法, 人称代名詞 (主格) の削除

Abstract

Quirk et al. (1985: 352) は、現代英語において、人称代名詞を修飾する際には厳しい制限があるが、いくつかのタイプは容認されると指摘している。しかし、Quirk et al. (1985: 352) の指摘がどのように我々に影響するのかを、より具体的に調査する必要がある。実際に、“Jack Sparrow, he of the gold teeth, hoop earring and (The Guardian, 2011/5/19)” などの事例は観察されるからである。そこで、本発表は、人称代名詞 (主格) が of 句により後置修飾される事例に焦点を絞り、現代英語における he/she of NP の特徴を提示する。まず、The Guardian (2011/5/19) の事例のように、固有名詞 (人名) の直後に、he/she of NP が同格用法として生起する事例については、COCA を用いた調査の結果、やや Newspaper のジャンルで用いられるという結果が得られた。さらに、

同格用法で用いられる he/she of NP の he/she が削除される事例を指摘し、その分析を行った (e.g., Jack Sparrow, of NP)。その理由は、人称代名詞の後置修飾を極力避けたいという発信者の意識によるものであると指摘した。そして、この言語事実は前述した Quirk et al. (1985: 352) の指摘と矛盾しないものである。

< Presentation Q&A session 6 >

【研究発表 1】

A Validation Study of the Accuracy of Lexical Diversity Tools

Hu XIAOLIN (Tokyo University of Foreign Studies)

Keywords: lexical diversity, tools evaluation, accuracy

Abstract

In recent years, a variety of lexical diversity tools are available for easy access to the analysis of a large amount of texts or corpora (Stuart Webb, 2019). However, few studies focus on the accuracy assessment for these tools.

This research aims to evaluate the accuracy of the lexical diversity tools and to explore the possible factors affecting the outputs of different tools. To this end, six widely used tools were tested on 24 transcriptions from NICT JLE Corpus (Izumi et al. 2004). The tools were divided into two groups according to their lemmatization principles. Group 1 (no lemmatization, all forms treated as different wordforms, i.e. types) contains Text Inspector (Bax, 2012), VocabProfile Program (Cobb, 2002), Coh-Metrix (Graesser, 2004) and CLAN (MacWhinney, 2000); Group 2 (with lemmatization, all inflected forms treated as one type) contains CLAN (MacWhinney, 2000), Lexical Complexity Analyzer (Lu, 2012) and TAALED (Kyle & Crossley, 2015). The type/token ratio (TTR) and Guiraud's index obtained by manually counted types and tokens were used as gold standards to evaluate the accuracy. This research only focuses on TTR and Guiraud, but the accurate counting and recognition of types and tokens will affect the other metrics as well.

The results show that in Group 1, Text Inspector has the most accurate mean scores in TTR against the gold standard compared with other tools (Error percentages: Text Inspector 0.33%, Coh-Metrix 1.00%, VocabProfile 2.22%, CLAN 4.20%). In CLAN(unlemmatized mode), a positive correlation between the proportion of contracted forms in texts and the error rates of TTR was observed ($r = 0.95$). For Group 2, CLAN(lemmatized mode) had the most robust performance in TTR (CLAN 0.70%, Lexical Complexity Analyzer 2.43%, TAALED 13.74%). Although a smaller difference was observed in Guiraud's index, Text Inspector (Group 1) and CLAN (Group 2) outperformed the other tools. Therefore, tools for lexical diversity should be selected carefully according to different data types and research aims.

【研究発表2】

A Corpus Based Approach to Creating an Advanced Wordbook for University Students

Satoru UCHIDA (Kyushu University)
Takehiko SHIMIZU (Kyushu University)
Saaya KIMURA (Kyushu University)

Keywords: advanced wordlist, vocabulary learning, university education

Abstract

The purpose of the present study is to report on the creation of an advanced vocabulary wordbook for university students. The book, titled “Word Quest” (Kyushu University Press), aims to cover advanced vocabulary that students are likely to encounter when developing their expertise. It contains about 1500 words and phrases in total with three main sections of “Academic”, “SDGs” (Sustainable Development Goals) and “Studying abroad.” The words in the former two sections are selected based on our original corpus consisting mainly of academic books and papers. Some examples are “agitate”, “ephemeral”, and “viable” from the academic section and “eradicate”, “infectious”, and “sewage” from the SDGs section. To the best of our knowledge, this is the first attempt to group words along the framework of SDGs, and it will be reported how we achieved this objective. It will be also shown how we created this book collaboratively with university students as an attempt to cover their actual needs in their current and future studies.

To examine the usefulness of our wordlist, a quantitative analysis was carried out using online news articles. It can be assumed that news passages contain advanced vocabularies that are essential for understanding the current issues around the world, hence important for university students who wish to get continuous updates on the world trend. For this purpose, 408 articles published in a week were collected from the BBC website with average length of 681.6 words. It will be shown that 70.6 percent of the articles contain at least three words from the book and 96.1 percent contains at least one. Even though the words in our list are mostly specialist terms and are not high in frequency, the results show that the book successfully covers what university students encounter when they research

【研究発表3】

付加疑問との連鎖関係からみた右方転位構造（テイル）の機能

山崎のぞみ（関西外国語大学）

Keywords: Spoken BNC2014, 右方転位構造（テイル）、付加疑問、発話末、連鎖

Abstract

本研究は、話し言葉に特徴的な右方転位構造—「テイル (tail)」と呼ぶ—が発話末で付加疑問 (TQ) と共起する場合、その連鎖関係がテイルの機能にどのように影響するかという点について分析する。

テイルは、節中の要素と同一指示の要素が再度、名詞句 (1) や代名詞 (2) の形式で発話末に現れる構造であり、しばしば付加疑問と共起する。

(1) so but it's a much better place now isn't it? The park? (Spoken BNC2014: SC7H)

(2) so it's a nice circuit that isn't it? (Spoken BNC2014: SGAN)

テイルには、指示対象の確認・明確化や強調機能、話者の評価や感情を聞き手と共有する対人関係機能があることが指摘されている (Aijmer, 1989; Biber et al., 1999; Durham, 2011; McCarthy & Carter, 1997; Rühlemann, 2007; Timmis, 2009, 2015)。

本調査では、Spoken BNC2014 から付加疑問と共起するテイルを収集した。TQ との連鎖順序とテイルの機能の関係を、発話が評価的か否かの別、テイルの構造や形式、さらにテイルの意味カテゴリーを組み込んで分析した。

その結果、名詞句テイルの場合、「TQ+ テイル」の順のテイルは指示対象を明確化・確認する機能が大きく、「テイル+ TQ」は、話者の態度表明や感情表出とより関わっていること示した。一方、代名詞テイルの場合、all of that など2語以上のテイルのほとんどが「TQ+ テイル」の順序で現れ、確認・強調機能との関連性を示したが、that など1語のテイルは「テイル+ TQ」の順序で現れる傾向が高かった。これらのことから、テイルが付加疑問と共起する時、共起する順序によって、テイルが果たす主な機能に違いが見られることや、名詞句テイルと代名詞テイルは、付加疑問との共起パターンが同じではないことが分かった。

【研究発表4】

経済学国際ジャーナルにおける Method の章のコーパス分析

中谷 安男 (法政大学経済学部)

Keywords: Academic Writing, Economics Journal Corpus, Research Methods, Applicability of Corpus Linguistics

Abstract

競争力の高い国際ジャーナルに論文が掲載されるためには、研究の信頼性と妥当性を確立し、論文の中で示した実験手法などを再現できるように客観的に書く必要がある (Cohen, 1994)。これらを明示するのがメソッド (Method) と呼ばれる研究計画や研究手法を記載する章である。査読者は、この章を精査し、研究手法が明確で、適切に結果を導きだしているか判断し採択を決める (中谷, 2016)。しかしながら Swales (2004) が指摘しているように、これまで Method の適切な書き方に関する論文は多くない。また、規模の大きなコーパスに基づき、信頼性の高い手法で特徴的な語彙やクラスター表現を分析した研究は少ない。特に、経済学の国際的論文のコー

パステータに基づく Method の検証はほとんどない (中谷, 2020)。

本論はこの点に注目して、インパクトファクターの高い以下の4つの学術雑誌から37本の論文を集め、約65万語の EJC (Economics Journal Corpus) を構築した。American Economic Review, Econometrica, Quarterly Journal of Economics, International Economic である。今回は、EJCの中から Method の章を抽出しコーパスを構築し、論文の他の章である Introduction, Result, Discussion の章のコーパスと比較検証した。AntConk Windows (3.5.8) を活用し2つのコーパスの Keyword 分析を行った。結果として Method における名詞や動詞の特徴語やクラスター表現を抽出できた。さらに、これらを活用した Move の構成が明らかになった。

< Symposium >

Symposium: 小・中・高における DDL 普及への挑戦 —DDL ツールの開発, 授業実践, 分野横断的考察—

西垣知佳子 (千葉大学), 赤瀬川史郎 (Lago 言語研究所)

水本 篤 (関西大学), 石井 雄隆 (千葉大学)

Peter Crosthwaite (University of Queensland), 安部 朋世 (千葉大学)

物井 尚子 (千葉大学), 小山 義徳 (千葉大学)

星野 由子 (千葉大学), 神谷 昇 (千葉大学)

Pichinart Kumpawan (Surasakmontree School)

Keywords: 発見学習, eDDL, hDDL, BES Search, 文法指導

Abstract

本シンポジウムは DDL SIG の企画である。DDL (data-driven learning) は、主に大学生を対象として世界的に利用が広がっている。本 SIG 企画は、日本の小・中・高校に DDL を導入するための試みと成果を、以下の4点から報告、検討する。

1) 小・中・高校生のための3種のウェブ DDL ツールと教材: eDDL, hDDL, BES Search

まず、これまでに学校現場で実施してきた DDL 実践の結果明らかになった DDL の成果を報告する。続いて、小・中・高校の英語授業で DDL を普及させるために開発した3種類のウェブ DDL ツールを紹介する。これらのツールは、学習者の英語力と認知レベルに合致する学習用コーパスと、使い易い検索ツールを搭載しており、小学生用 DDL ツール (eDDL), ならびに中・高生用 DDL ツール (hDDL) を開発した。併せて、教師が DDL 教材作成に利用できる入門・初級レベルの英文を収集した検索ツール (BES Search) を開発した。全て登録不要、無料で利用できる。

2) DDL 実践の分野横断的考察

教育心理学の観点から、DDL における発見学習の効果を検討する。

また、発表者らの行う DDL では、学習者は帰納的に英語の文法規則を発見し、日本語で言語化する。学習者の発見内容を見てみると、国語科で学んだ知識を使って、英語の文法規則を記述していることがわかった。そこで、文法に関連する事柄が、

外国語科と国語科の検定教科書で、どのようなことが、どのように学ばれているかを調査した。その結果を報告する。

3) DDL の指導と評価

既存の大学生用 DDL 評価テストと質問紙調査を検討し、小・中・高校生用の評価テストと質問紙の開発に向けた構想を述べる。

4) 国際的視点から考える DDL

本プロジェクト独自の DDL をタイ・バンコクの中学校英語教師が実践した事例を紹介する。また、pre-tertiary learners の DDL 研究の先駆者である University of Queensland の Crosthwaite 先生に、海外での初等・中等教育における DDL についてお話しをいただく。

以上を踏まえて、最後に、日本の初等・中等教育における DDL の活用について検討する。

Timetable of symposium

Part 1: Data-driven learning for younger learners: Current issues, future directions.

Peter Crosthwaite (University of Queensland)

Part 2: 小・中・高校生のための3種のウェブ DDL ツールと教材：eDDL, hDDL, BES Search

西垣知佳子 (千葉大学), 赤瀬川史郎 (Lago 言語研究所),
石井 雄隆 (千葉大学), 神谷 昇 (千葉大学). 西垣が発表

Part 3: 小・中・高校における DDL 実践の分野横断的考察

- | | |
|---------------------|---------------------|
| ①発見学習から DDL の効果を考える | 小山 義徳 (千葉大学) |
| ②小学校英語教育の分野から | 物井 尚子, 星野 由子 (千葉大学) |
| ③小学校国語科教育の分野から | 安部 朋世 (千葉大学) |

Part 4: DDL の指導と評価

水本 篤 (関西大学)

Part 5: Communicative English Classroom using DDL in Thailand

Pichinart Kumpawan (Surasakmontree School)

『英語コーパス研究』 投稿規定

(2019年10月改定)

1. 投稿資格

投稿は会員に限る。共著の場合、第一著者は会員であることとし、その他の共著者については会員でなくてもよい。

2. 原稿の種類と長さ

【研究論文】

英文 A4サイズ 1ページあたり35行（文字数の指定はしない）、周囲の余白1インチ（25.4mm）、17ページ以内（Times New Roman 10.5ポイント使用）

和文 A4サイズ 1ページあたり35行（文字数の指定はしない）、投稿時17枚以内（明朝体フォント（游明朝・ヒラギノ明朝など）10.5ポイント使用）

※和文中の英文のフォントについては Times New Roman を原則とする。Century は用いてはならない。

（いずれも Abstract（英文300語以内）、図表、注、参考文献目録、付録、謝辞、著者情報などを含む。）

【研究ノート、総説論文・書評論文（Review article, Book review）】

- ・研究ノート：論文のカテゴリーに属さない小論文や萌芽的な研究、新しい研究開発の成果などをまとめたもの

- ・総説論文：体系的かつ網羅的に先行研究をまとめたもの

- ・書評論文：専門書の研究分野への貢献と課題点を明確にしたもの

英文 A4サイズ 1ページあたり35行（文字数の指定はしない）、周囲の余白1インチ（25.4mm）、12ページ以内（Times New Roman 10.5ポイント使用）

和文 A4サイズ 1ページあたり35行、投稿時12枚以内（明朝体フォント（游明朝・ヒラギノ明朝など）10.5ポイント使用）

※和文中の英文のフォントについては Times New Roman を原則とする。Century は用いてはならない。

（いずれも Abstract（英文300語以内）、図表、注、参考文献目録、付録、謝辞、著者情報などを含む。）

【その他（ソフトウェアレビュー、書評（図書紹介）、コーパス紹介など）】

研究論文の半分以内の分量

3. 原稿作成時の注意

下記のように投稿者を特定できるような情報、その他、本人の同定につながると考えられる情報は、採用決定後の最終原稿に追記するものとし、投稿時には記載しないこと。

- (1) 謝辞など
- (2) 「本論は、英語コーパス学会第X回大会において口頭発表した内容に加筆修正を施したものである。」などの文言
- (3) 「筆者が収集し、WWW (<http://...>) で公開しているデータ…」など、筆者情報につながる URL 情報など
- (4) 「拙論 (2006) で論じたように…」などと記して、参考文献目録で当該文献を参照している場合、「拙論」ではなく著者 (2006) として表記すること。

4. 提出方法など

- (1) 下記の (A) 原稿ファイル (Microsoft Word で作成したファイルとその PDF ファイル)、(B) 著者情報ファイル、(C) 論文投稿チェックシートの3種類のファイルを電子メール添付で提出。(B)、(C) については Web 掲載のフォーマットを使用のこと。
- (2) 電子メールの件名 (Subject) は「『英語コーパス研究』投稿原稿 (著者氏名)」とすること。
- (3) 提出先、締め切り期日等に関しては学会 Web サイトを参照のこと。

(A) 原稿ファイル

- a. 提出するファイル名は「原稿題名 (著者氏名)」とすること。
- b. 原稿題名の前に「論文」、「研究ノート」、「総説論文」、「書評論文」、「コーパス紹介」などの種類を明記すること。
- c. 原稿本体の冒頭には上記種類の別と題名のみを記すこと。

(B) 著者情報ファイル：「著者情報 (著者氏名)」

- a. 和文原稿の場合は英文タイトル、英文原稿には和文タイトル
- b. 著者氏名 (ふりがな・ローマ字表記)
- c. 所属
- d. 郵便番号・住所・電話番号
- e. 電子メールアドレス

(C) 論文投稿チェックシート：「論文投稿チェックシート (著者氏名)」

Web 掲載のチェックシートの必要項目すべてに☑を入れること。

5. スタイル

投稿論文は、研究論文、研究ノート、総説論文・書評論文の別、また、和文・英文の別にかかわらず、『英語コーパス研究』スタイルシートに従い執筆することとする。

6. 掲載論文等の電子化

掲載された論文等の著者は、論文等を電子化して学会ホームページで公開することに同意する。

7. 著作権

掲載された論文等の著作権は、本学会に帰属する。本学会は掲載論文等を印刷媒体・電子媒体で公開する権利を有するものとする。ただし、著者が自著論文等を自分のホームページに掲載したり、自著の本に転載したりすることは妨げない。

8. 研究倫理

投稿にあたっては、下記文書などを参照し、不正行為のないようにすること。
独立行政法人科学技術振興機構『研究者のみなさまへ～研究活動における不正行為の防止について～』<https://www.jst.go.jp/contract/kisoken/h25/others/h25s805others131120.pdf>

英語コーパス学会会則

(名称)

第1条 本会は「英語コーパス学会」(Japan Association for English Corpus Studies, 略称 JAECS) と称する。

(目的)

第2条 本会は英語コーパス及びコーパスツールの開発・評価・利用に関わる研究、また、英語コーパスを用いた言語研究・言語教育研究・関連研究を促進することを目的とする。

(事業)

第3条 本会は前条の目的を達成するために、次の事業を行う。

- (1) 大会・研究会等の開催
- (2) 学会誌・学会報等の発行
- (3) その他本会の趣旨に沿う事業

(会員)

第4条 本会の会員は一般会員、学生会員、団体会員、賛助会員、功勞会員及び名誉会員よりなる。

- (1) 一般会員は本会の趣旨に賛同する個人とする。
- (2) 学生会員は本会の趣旨に賛同する個人のうち、大学又は大学院に籍を置く学生とする。
- (3) 団体会員は本会の趣旨に賛同する大学、研究所、図書館その他の研究・教育団体とする。
- (4) 賛助会員は本会の趣旨に賛同する企業等とする。
- (5) 功勞会員は本会の活動に長く寄与した個人とする。功勞会員の規程は別に定める。
- (6) 名誉会員は本会の活動に特別に寄与した個人とする。

(会費)

第5条 本会の会費について以下の通り定める。

- (1) 会員は所定の会費を納めるものとする。
- (2) 会費の額については次の通りとする。

一般会員	年額	5,000円	(在外会員は年額12,000円)
学生会員	年額	2,000円	(在外会員は年額10,000円)
団体会員	年額	5,000円	
賛助会員	年額	15,000円	
- (3) 会費は入会時点又は会計年度開始時点で納入する。
- (4) 2年間にわたって会費納入がない場合は会員の資格を失う。

- (5) 名誉会員，功勞会員，顧問からは会費を徴収しない。

(会計年度)

第6条 本会の会計年度は4月1日に始まり，翌年3月31日をもって終わる。

(組織)

第7条 本会に執行部，事務局，役員会，学会誌編集委員会，学会賞選考委員会，大会実行委員会，研究会（SIG）を置く。

- (1) 執行部は会長，副会長，事務局長，事務局員で構成し，本会全体にかかわる事業を執行・監督する。
- (2) 事務局は事務局長及び事務局員で構成し，本会の事務を執行する。
- (3) 役員会は役員で構成し，本会にかかる諸問題を審議・決定する。
- (4) 学会誌編集委員会は学会誌の刊行にかかる業務を担当する。学会誌編集委員会の規程は別に定める。
- (5) 学会賞選考委員会は学会賞・奨励賞の選考にかかる業務を担当する。学会賞選考委員会の規程は別に定める。
- (6) 大会実行委員会は大会の企画・準備・実施にかかる業務を担当する。大会実行委員会の規程は別に定める。
- (7) 研究会（SIG）は会員のうち，希望する者によって構成し，それぞれが掲げる研究目的に応じた活動を行う。研究会の規程は別に定める。

(役員)

第8条 本会に次の役員をおく。

- (1) 会 長 1名
- (2) 副会長 若干名
- (3) 理 事 若干名
- (4) 幹 事 若干名

(役員任期・定年)

第9条 役員任期は以下の通りとする。

- (1) 会長・副会長の任期は2年とし，引き続き2期までの再任を妨げない。
- (2) 理事・幹事の任期は2年とし，再任を妨げない。
- (3) 任期は当該年度の4月1日から起算する。
- (4) 役員定年を70歳とする。任期途中で定年に達したときは当該年度の終了まで，その任にあたる。

(役員任務)

第10条 役員任務は以下の通りとする。

- (1) 会長は本会を代表し，会務を統括する。会長は総会・役員会を招集し，これを主宰する。

- (2) 副会長は会長の命ずる職務を所掌するとともに、会長を補佐し、必要に応じて会長の職務を代行する。
- (3) 理事は役員会に出席し、本会の運営に関わる重要事項を審議・議決する。
- (4) 幹事は役員会に出席し、理事を補佐し、本会の運営に関わる重要事項を審議・議決する。

(役員を選出)

第11条 役員は役員会における投票によって決定する。

(役職員)

第12条 本会に次の役職員をおく。

- | | |
|-----------------|-----|
| (1) 顧問 | 若干名 |
| (2) 事務局長 | 1名 |
| (3) 事務局員 | 若干名 |
| (4) 監査 | 1名 |
| (5) 学会誌編集委員会委員長 | 1名 |
| (6) 学会誌編集委員 | 若干名 |
| (7) 学会賞選考委員会委員長 | 1名 |
| (8) 学会賞選考委員 | 若干名 |
| (9) 大会実行委員会委員長 | 1名 |
| (10) 大会実行委員 | 若干名 |

(役職員の任期・定年)

第13条 役職員の任期は以下の通りとする。

- (1) 顧問の任期は終身とする。
- (2) 事務局長・事務局員、監査、学会誌編集委員会委員長及び委員、学会賞選考委員会委員長及び委員の任期は2年とし、引き続き2期までの再任を妨げない。任期は当該年度の4月1日から起算する。
- (3) 大会実行委員会委員長及び委員の任期は、役員会で承認された日から当該大会に関係する業務の終了時までとする。
- (4) 顧問を除く役職員の定年を70歳とする。任期の途中で定年に達したときは当該年度の終了まで、その任にあたる。

(役職員の任務)

第14条 役職員の任務は以下の通りとする。

- (1) 顧問は役員会の求めに応じて学会運営への助言を行う。
- (2) 事務局長は事務局を主宰し、本会の事務を執行・監督する。
- (3) 事務局員は事務局長の指示の下、必要な業務を執行する。
- (4) 監査は本会の会計及び運営が適切になされているか精査し、その結果を総会で報告する。
- (5) 学会誌編集委員会委員長は学会誌編集委員会を主宰し、学会誌の刊行

にかかる業務を執行・監督する。

- (6) 学会誌編集委員は委員長の指示の下、必要な業務を執行する。
- (7) 学会賞選考委員会委員長は学会賞選考委員会を主宰し、学会賞・奨励賞の選考にかかる業務を執行・監督する。
- (8) 学会賞選考委員は委員長の指示の下、必要な業務を執行する。
- (9) 大会実行委員会委員長は大会実行委員会を主宰し、大会の企画・準備・実施にかかる業務を執行・監督する。
- (10) 大会実行委員は委員長の指示の下、必要な業務を執行する。

(役職員の選出)

第15条 役職員は会長が推薦し、役員会で承認する。役職員と役員の兼務を妨げない。

(会議)

第16条 本会は以下の会議を開催する。

- (1) 総会は会長の招集により、原則として年1回以上開催し、会則の改定、予算・決算その他重要事項を審議する。なお、電子メールやその他の手段を用いた総会の開催も可能とする。総会での議決は出席者の過半数による。
- (2) 役員会は会長の招集により、原則として年2回以上開催し、本会の運営にかかる諸問題を審議し、決定する。なお、電子メールやその他の手段を用いた役員会の開催も可能とする。役員会での議決は出席者の過半数による。
- (3) 事務局会議は事務局長の判断の下、不定期に開催する。
- (4) 学会誌編集委員会、学会賞選考委員会、大会実行委員会は各委員長の判断の下、不定期に開催する。

付則

- (1) 本会則は2020年4月1日から施行する。
- (2) 本会則は2021年4月1日から改正施行する。

(備考)

- (1) 本会は1993年4月1日に「英語コーパス研究会」として発足し、1997年4月1日に「英語コーパス学会」に改組されて現在に至る。
- (2) 本会の事務局を九州大学言語文化研究院内田論研究室（〒819-0395 福岡市西区元岡744）に置く。

英語コーパス研究 (第28号)
【2021年5月31日発行】

編集・発行 ©2021 英語コーパス学会
〒819-0395 福岡県福岡市西区元岡744
九州大学大学院言語文化研究院 内田諭研究室気付
E-mail (事務局長) : jaecs.hq@gmail.com
Website: <http://jaecs.com/>
郵便振替口座 : 009300-3-195373 (英語コーパス学会)

印刷所 株式会社ニシキプリント
〒739-2117 東広島市高屋台2-1-12

English Corpus Studies: Vol.28 2021

Articles

Yasunori NISHINA / Corpus-Assisted Discourse Studies in Airline Company Profiles Through the Lens of Moves and Adjectives	1
Nozomi YAMASAKI / Exploring Functions of Tails from the Viewpoint of Their Co-occurrence with Tag Questions.....	27
Conference Program & Abstract	
The 46th Conference of Japan Association for English Corpus Studies	47

JAPAN ASSOCIATION FOR ENGLISH CORPUS STUDIES

